

ROBUSTIFIZIERUNG VON SCHÄTZERN DURCH WINSORISIEREN AN ELLIPSOIDEN

DISSERTATION

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Fakultät für Mathematik
der Ruhr-Universität Bochum

von

WALTER OLBRICHT

BOCHUM 1988

ROBUSTIFIZIERUNG VON SCHÄTZERN DURCH
WINSORISIEREN AN ELLIPSOIDEN

Digitized by the Internet Archive
in 2026



ROBUSTIFIZIERUNG VON SCHÄTZERN DURCH WINSORISIEREN AN ELLIPSOIDEN

DISSERTATION

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Fakultät für Mathematik
der Ruhr-Universität Bochum

von

WALTER OLBRICHT

BOCHUM 1988

Dissertation eingereicht am: 22. 3. 1988

Tag der mündlichen Prüfung: 13. 7. 1988

Dekan: Prof. Dr. A. T. Huckleberry

Gutachter: Prof. Dr. V. Baumann

Prof. Dr. S. Albeverio

Prof. Dr. P. J. Huber (Harvard University)

INHALTSVERZEICHNIS

<u>0. EINLEITUNG</u>	1
<u>1. ROBUSTIFIZIEREN DURCH WINSORISIEREN AN ELLIPSOIDEN</u>	
1.1. Hubers Lokationsschätzer als MK-Schätzer	5
1.2. Zur Definition robustifizierter Schätzer	8
1.3. Allgemeine Aussagen über das Schätzprinzip	12
<u>2. LOKATIONS- UND SKALENSCHÄTZUNG IM EINDIMENSIONALEN FALL</u>	
2.1. Schätzung eines eindimensionalen Parameters	22
2.2. Schätzung eines Lokationsparameters	25
2.3. Schätzung eines Skalenparameters	29
2.4. Schätzung eines Lokations- und Skalenparameters	32
2.5. Anmerkungen zur Wahl des Schätzers	54
2.6. Robustheit und Information	63
<u>3. LOKATIONS- UND SKALENSCHÄTZUNG IM MEHRDIMENSIONALEN FALL</u>	
3.1. Allgemeines	72
3.2. Teilaussagen im zweidimensionalen Fall	76
3.3. Teilaussagen im mehrdimensionalen Fall	83
3.4. Vollständige Aussagen unter abgeschwächten Bedingungen	86
<u>4. WEITERE ANWENDUNGSMÖGLICHKEITEN</u>	
4.1. Robustifizierung einiger klassischer Modelle	90
4.2. Indirekte Robustifizierung	92
4.3. Robustifizierung von Varianzkomponentenmodellen	95
<u>5. SCHLUSSBEMERKUNGEN</u>	112
<u>LITERATURVERZEICHNIS</u>	114

0. EINLEITUNG

Statistische Modelle sind stets Idealisierungen der Wirklichkeit; die ihnen zugrundeliegenden Annahmen treffen daher im allgemeinen nicht exakt, sondern nur annähernd zu. Dementsprechend müssen brauchbare statistische Verfahren gegenüber kleinen Abweichungen von den theoretischen Annahmen "unempfindlich" sein. Die vorliegende Arbeit befaßt sich mit der Entwicklung solcher robuster Verfahren. Die Ausgangslage und die Zielsetzung lassen sich - notwendigerweise sehr skizzenhaft - mit den folgenden Thesen umreißen:

- (i) Es wird davon ausgegangen, daß Ausreißer und Modellabweichungen ein Faktum der statistischen Praxis sind - auch wenn es nicht oder nicht leicht möglich ist, dies präzise zu definieren. Die Notwendigkeit robuster Verfahren steht von daher außer Frage.
- (ii) Die Möglichkeit, robuste Verfahren zu entwickeln, scheint durch die Entwicklung der letzten zwanzig Jahre ebenfalls erwiesen zu sein. Dies gilt insbesondere für die von P.J. Huber und F.R. Hampel vertretene Auffassung, nach der robuste Statistik zur parametrischen Statistik gehört und versucht, Verfahren gegen "kleine" Abweichungen von den idealen (parametrischen) Modellannahmen unempfindlich zu machen.
- (iii) "Robustheit" ist keine Eigenschaft des zu schätzenden Funktionals, sondern allein der Schätzerfolge. Das Symmetriezentrum einer eindimensionalen symmetrischen Verteilung kann etwa sehr robust (z.B. durch den Median) oder sehr wenig robust (z.B. durch das arithmetische Mittel) geschätzt werden.
- (iv) Die weitaus meisten Untersuchungen beziehen sich bislang auf den Fall robuster Lokationsschätzungen im eindimensionalen i.i.d. Fall. Skalenschätzungen werden meist nur als "nuisance parameter" mitbetrachtet. Relativ ausführlich

wurde auch Regression untersucht. Bei der Schätzung von multivariater Lokation und Scatter gibt es demgegenüber schon i.i.d. Fall wenig gesicherte Erkenntnisse. Untersuchungen für speziellere Modelle (z.B. Varianzkomponentenmodelle) fehlen von einigen ersten Ansätzen abgesehen völlig. Gerade in solchen "unübersichtlichen" Modellen ist es jedoch besonders schwierig, "Ausreißer" direkt zu erkennen.

- (v) Die Parameter genetischer Modelle besitzen häufig eine substanzwissenschaftliche Interpretation (z.B. Heritabilitätskoeffizienten), auf die aus inhaltlichen Gründen nicht verzichtet werden kann. Der etwa von Bickel/Lehmann (1975) vorgeschlagene Weg, irgendein "brauchbares" (robust schätzbares) Funktional zu wählen, ist daher hier nicht begehbar. In Verbindung mit (iii) liegt vielmehr der Gedanke nahe, schon vorhandene Schätzer zu "robustifizieren", d.h. gegen Ausreißer "unempfindlicher" zu machen. Von den naheliegenden Möglichkeiten dazu scheint aus folgenden Gründen Winsorisierung besonders vielversprechend:
 - a) Dieses Prinzip hat sich in den bisherigen Robustheitsstudien sehr gut bewährt (Hubers Proposal 2, "Huberizing").
 - b) Gerade bei genetischen Fragestellungen liegen häufig nur wenige Daten vor, so daß sich "Trimming" nicht empfiehlt.
 - c) Schließlich kann auch bei Ausreißern die "Richtung des Ausreißers" noch wichtige Information enthalten.
- (vi) Im eindimensionalen Lokationsfall gibt es eine verwirrende Vielfalt von Schätzvarianten, die häufig noch durch "tuning constants" auf die verschiedensten Fälle optimal eingestellt werden können, was aber einerseits ein gewisses "Vorwissen" erfordert und andererseits ihre Interpretation erschwert. In komplizierteren Modellen wird (noch weniger als im eindimensionalen Lokationsfall) in der Regel ein derartiges Wissen nicht zur Verfügung stehen.
- (vii) Aus den obigen Anmerkungen ergeben sich die folgenden Ziele:
 - a) Es werden Methoden gesucht, die u.a. eine robuste Behandlung von Modellen mit komplizierter Kovarianzstruktur ermöglichen sollen.
 - b) Es geht um möglichst durchsichtige, wenig willkürliche Verfahren mit breitem Anwendungsbereich. Es wird versucht, dies durch Robustifizieren bekannter Schätzer

mittels Winsorisierung zu erreichen.

- c) Die vorgeschlagenen Methoden müssen durch Vergleich mit gegebenenfalls schon vorhandenen Verfahren abgesichert werden. Dies kann durch theoretische Überlegungen (Optimalitätseigenschaften, asymptotische Effizienz), Simulation und (soweit möglich) Anwendung auf konkrete Daten geschehen.

Die Arbeit ist folgendermaßen aufgebaut:

Im ersten Kapitel wird das Robustifizieren durch Winsorisieren an Ellipsoiden zunächst heuristisch motiviert, dann allgemein definiert und schließlich im allgemeinen Rahmen untersucht. Die Fragestellung wird dabei technisch gesehen als Fixpunktproblem oder Nullstellenproblem behandelt. Es werden Hilfsmittel bereitgestellt, die zum Nachweis der Existenz und Eindeutigkeit von robusten Schätzungen, zur Konstruktion von Algorithmen zur Berechnung sowie zur Analyse von statistischen Eigenschaften benutzt werden können. Vom üblichen Vorgehen weicht die Darstellung vor allem in vier Punkten ab:

1. Es wird konsequent von der Robustifizierung gegebener (statt der Herleitung neuer robuster) Schätzer ausgegangen.
2. Die Robustifizierung geschieht durch Winsorisierung am Konzentrationsellipsoid - also einem durch das zugrundeliegende Modell häufig fast "kanonisch" determinierten Objekt der deskriptiven Statistik.
3. Der Eindeutigkeitsnachweis wird auf eine Homotopieüberlegung zurückgeführt; dies liefert zugleich einen Ausgangspunkt für brauchbare Algorithmen (Path-following Methoden).
4. Durch Benutzung von nonsmooth Analysis können Maße mit und ohne Atomen einheitlich behandelt werden.

Das zweite Kapitel behandelt ausführlich die Robustifizierung der Standardschätzer Mittel und Varianz im eindimensionalen Fall. Nach dem einführenden Abschnitt 2.1. werden in Abschnitt 2.2. die Lokationsschätzung, in Abschnitt 2.3. die Skalenschätzung und in Abschnitt 2.4. die simultane Lokations- und Skalenschätzung untersucht. Es zeigt sich, daß die im ersten Kapitel bereitgestellten Konzepte dem Problem sehr angemessen sind. Weiterhin wird die enge Beziehung zum Momentenproblem herausgearbeitet und die Robustifizierung in der vorliegenden Situation im wesentlichen als "generalisiertes Momentenproblem" interpretiert. Dies liefert nebenbei

auch eine strukturell interessante Erweiterung bisheriger Eindeutigkeitsresultate. In den Abschnitten 2.5. und 2.6. wird mittels datenanalytischer und informationstheoretischer Überlegungen auf die Realitätsnähe und praktische Brauchbarkeit eingegangen.

Das dritte Kapitel konzentriert sich vor allem auf die völlig offene Frage der Eindeutigkeit robuster Schätzungen im mehrdimensionalen Fall. Einige Teilresultate hierzu sowie einige Alternativvorschläge können angegeben werden.

Im vierten Kapitel werden einige klassische Modelle robustifiziert und Anwendungen in der Statistischen Genetik sowie der Meßfehleranalyse aufgezeigt. Eine Monte-Carlo-Studie zur robusten Schätzung des Innerklassenkorrelationskoeffizienten in einer balanzierten Ein-Weg-Klassifikation gibt einen Eindruck von den quantitativen Aspekten der Robustifizierung.

Das fünfte Kapitel enthält einige abschließende Bemerkungen zu weiterführenden Fragen.

• • •

Der Autor möchte an dieser Stelle Herrn Professor Dr. V. Baumann für die anregende und hilfreiche Betreuung dieser Arbeit herzlich danken. Frau M. Howahl und Frau E. Panten dankt er für die Anfertigung des größten Teils der Reinschrift.

*Gleich an die Grenzen der Wissenschaft zu gehen.
Es läßt sich bald lernen, wo es noch fehlt.*

(G. Chr. Lichtenberg)

1. ROBUSTIFIZIEREN DURCH WINSORISIEREN AN ELLIPSOIDEN

In diesem Kapitel soll ein allgemeines Verfahren beschrieben werden, mit dem man in vielen Situationen gegebene Schätzer robustifizieren kann. Das Verfahren wird zunächst heuristisch am Beispiel des klassischen "Huber-Schätzers" (für einen eindimensionalen Lokationsparameter) dargestellt (Abschnitt 1.1.), dann allgemein definiert (Abschnitt 1.2.) und schließlich im allgemeinen Rahmen untersucht (Abschnitt 1.3.). Abschnitt 1.3. stellt vorwiegend Hilfsmittel bereit, die in spezielleren Situationen zur Untersuchung der Eigenschaften genutzt werden können.

1.1. Hubers Lokationsschätzer als MK-Schätzer

In seiner grundlegenden Arbeit (Huber 1964) schlägt Huber u.a. als Lokationsschätzer (bei bekanntem Skalenwert) auf der Basis einer Stichprobe $x_1, \dots, x_n$ die Minimalstelle der Funktion

$$t \mapsto \sum_{i=1}^n \rho(x_i - t) \quad (1)$$

vor, wobei

$$\rho(x) = \begin{cases} \frac{1}{2}x^2 & , \quad |x| \leq k \\ k|x| - \frac{1}{2}k^2 & , \quad |x| > k \end{cases} \quad (2)$$

ist. Er weist bereits darauf hin, daß dieser Schätzer ein winsorisiertes Mittel der Stichprobe ist (Huber 1964, S. 75). Dabei handelt es sich genauer um ein "metrisch winsorisiertes" Mittel im Unterschied zum rein datenanalytisch definierten " α -winsorisierten" Mittel im Sinne von Tukey (vgl. auch Bickel 1965, S. 848). Trotz dieser anschaulichen Darstellung im Stichprobenfall gelten die Huber-Schätzer wegen einer fehlenden expliziten Darstellung als intuitiv schwer erfassbar (Lehmann

1983, S. 56f). Im folgenden soll deshalb zunächst untersucht werden, inwieweit der Huber-Schätzer als Minimum-Kontrast-Schätzer im Sinne der folgenden Definition interpretiert werden kann:

1.1.1. Definition (Minimum-Kontrast-Schätzer): Es sei $\mathcal{P}$ eine Familie von W-Maßen auf einem Meßraum $(X, \mathcal{A})$, $(T, \mathcal{U})$ ein topologischer Raum, $\tau: \mathcal{P} \rightarrow T$ eine Abbildung ("Parametrisierung von $\mathcal{P}$ ") und $\sigma(\mathcal{U})$ die Borel- σ -Algebra auf T .

- (a) Eine Familie $\mathcal{A}$ -meßbarer Funktionen $f_t: X \rightarrow [-\infty, \infty]$, $t \in T$, heißt eine Familie von Kontrastfunktionen für $\mathcal{P}$, falls
- (i) für alle $P \in \mathcal{P}$ und alle $t \in T$ $E_P(f_t)$ existiert
 - (ii) für alle $P \in \mathcal{P}$ und alle $t \neq \tau(P)$ gilt:

$$E_P(f_{\tau(P)}) < E_P(f_t). \quad (3)$$

- (b) Eine Minimum-Kontrast-Schätzung zum Stichprobenumfang n ist eine $(\mathcal{A}^n, \sigma(\mathcal{U}))$ -meßbare Abbildung $\varphi_n: X^n \rightarrow \tau(\mathcal{P})$ mit

$$\frac{1}{n} \sum_{i=1}^n f_{\varphi_n(x_1, \dots, x_n)}(x_i) = \inf \left\{ \frac{1}{n} \sum_{i=1}^n f_t(x_i) : t \in \tau(\mathcal{P}) \right\}. \quad (4)$$

1.1.2. Notiz: Die obige Definition stammt aus Pfanzagl (1969, S. 251), wo sie allerdings noch etwas allgemeiner gefaßt wird. Die Minimum-Kontrast-Methode ist ein sehr allgemeines Verfahren zur Konstruktion "schnell konsistenter Schätzfolgen" (s. hierzu Strasser 1985, S.46ff).

1.1.3. Fragestellung: Für welche Menge von W-Maßen $\mathcal{P}$ (auf $(\mathbb{R}, \mathcal{B})$) und welche meßbaren Funktionen $\tau: \mathcal{P} \rightarrow \mathbb{R}$ gilt:

$$\int \rho(x - \tau(P)) dP(x) < \int \rho(x - t) dP(x) \quad \forall t \neq \tau(P) \quad \forall P \in \mathcal{P} ? \quad (5)$$

1.1.4. Anmerkung: Setzt man für ein gegebenes W-Maß auf $(\mathbb{R}, \mathcal{B})$

$$Q(t) := \int \rho(x - t) dP(x), \quad (6)$$

so gelten die Äquivalenzen

$$\forall t \in \mathbb{R} : Q(t) < \infty \iff \exists t_0 \in \mathbb{R} : Q(t_0) < \infty \iff E_P(|X|) < \infty.$$

Man kann (bzw. muß) sich daher auf W-Maße mit finitem ersten Moment beschränken. Aus diesem Grunde lehnt Huber (1981, S. 43) eine "funktionale" Definition seines Schätzers ab (s. auch Pfanzagl 1981, S. 80f).

1.1.5. Satz: Es sei P ein W -Maß auf $(\mathbb{R}, \mathcal{L})$ mit $E_P(|X|) < \infty$. Dann gilt:

(i) Die Funktion Q aus (6) ist auf $\mathbb{R}$ differenzierbar, und es ist

$$Q'(t) = -\int \psi(x - t) dP(x), \quad (7)$$

wobei

$$\psi(x) = \rho'(x) = \begin{cases} -k & , \quad x < -k \\ x & , \quad -k \leq x \leq k \\ k & , \quad x > k \end{cases} \quad (8)$$

ist.

(ii) Die Funktion Q' aus (7) ist auf $\mathbb{R}$ stetig und schwach monoton steigend. Ferner ist $\lim_{t \rightarrow -\infty} Q'(t) = -k$ und $\lim_{t \rightarrow \infty} Q'(t) = k$.

(iii) Q ist schwach konvex.

Beweis: (i) folgt durch Differentiation unter den Integralzeichen; (ii) und (iii) sind klar (vgl. auch Huber 1964, S. 76, für den Stichprobenfall).

Mit Blick auf die Fragestellung aus 1.1.3. ergibt sich also folgendes Resultat:

1.1.6. Satz: Es sei P ein W -Maß mit $E_P(|X|) < \infty$. Dann gilt:

(i) Ist $Q(t_0) < Q(t)$ für ein festes t_0 und alle $t \neq t_0$, so folgt

$$t_0 = \int_{[t_0-k, t_0+k]} x \, dP(x) + (t_0-k)P(]-\infty, t_0-k]) + (t_0+k)P(]t_0+k, \infty[). \quad (9)$$

(ii) Die Menge der Stellen $t_0 \in \mathbb{R}$ mit (9) ist ein kompaktes Intervall $K(P) = [c_1, c_2]$ (mit $c_1 \leq c_2$).

(iii) Jedes $t_0 \in K(P)$ ist globale Minimalstelle im schwachen Sinne von Q .

(iv) Genau dann, wenn $K(P) = [c_1, c_2] = \{c\}$ aus genau einem Punkt besteht, existiert ein $\tau(P)$ - nämlich c - mit

$$Q(\tau(P)) < Q(t) \quad \forall t \neq \tau(P).$$

(v) $K(P)$ enthält genau dann ein nichtentartetes Intervall $[a, b]$ ($a < b$), wenn gilt: $P(]-\infty, a-k]) = \frac{1}{2} = P(]b+k, \infty[)$.

Beweis: Die Aussagen ergeben sich unmittelbar.

1.1.7. Bemerkung:

(i) Teil (v) von Satz 1.1.6. charakterisiert die W -Maße (mit endlichem absoluten ersten Moment), für die der

Huber-Schätzer nicht als Minimum-Kontrast-Schätzer interpretiert werden kann. Das sind diejenigen Maße, für die der Median nur bis auf ein Intervall der Länge $\ell > 2k$ bestimmt ist.

- (ii) Ist P ein W-Maß, für das der Erwartungswert $m(P)$, der Median $\text{Med}(P)$ und der Huber-Schätzwert $H(P)$ eindeutig bestimmt sind, so läßt sich über den Zusammenhang von $m(P)$, $\text{Med}(P)$ und $H(P)$ nur die Aussage $H(P) \in [\text{Med}(P) - k, \text{Med}(P) + k]$ machen.
- (iii) Definiert man für jedes $t \in \mathbb{R}$ das "auf das Ellipsoid $[t-k, t+k]$ konzentrierte" W-Maß P_t durch

$$P_t(A) = P(A \cap [t-k, t+k]) + \frac{1}{A} (t-k) \cdot P([-\infty, t-k]) + \frac{1}{A} (t+k) \cdot P([t+k, \infty]),$$

so läßt sich der Huber-Schätzer $t_0 = H(P)$ auf folgende Arten beschreiben:

- a) t_0 ist Fixpunkt der Abbildung $t \mapsto \int x \, dP_t(x)$ ("Fixpunktproblem").
- b) t_0 ist Nullstelle der Abbildung $t \mapsto \int \psi(x-t) \, dP(x)$ ("generalisiertes Momentenproblem").
- c) t_0 ist die Stelle, an der das Mittel der konzentrierten Verteilung P_t mit dem Mittelpunkt des Intervalls $[t-k, t+k]$ übereinstimmt. Da der Intervallmittelpunkt zugleich Mittel jeder um t symmetrischen Verteilung ist, kann man dies auch so ausdrücken: Man paßt eine symmetrische Verteilung dadurch an P an, daß für die auf das Ellipsoid $[t-k, t+k]$ konzentrierten Verteilungen die ersten Momente übereinstimmen ("Approximationsproblem").

Die drei Sichtweisen sind eng miteinander verwandt; sie dienen im folgenden als Ausgangspunkt für eine Verallgemeinerung.

1.2. Zur Definition robustifizierter Schätzer

1.2.1. Allgemeine Beschreibung eines Schätzprinzips:

Es sei gegeben:

- (1) ein Banachraum $\mathcal{X}$ (= vollständiger normierter Vektorraum über

$\mathbb{R}$) (Die Norm in $\mathcal{X}$ sei mit $\|\cdot\|$ bezeichnet; $\mathcal{B}$ sei die σ -Algebra der Borel-Mengen.)

(2) eine Familie $\mathcal{W}$ von W -Maßen auf $(\mathcal{X}, \mathcal{B})$

(3) ein meßbarer Raum $(\Gamma, \mathcal{G})$ und eine Abbildung $\gamma : \mathcal{W} \longrightarrow \Gamma$

(4) eine Schätzerfolge d_n für $\gamma(P)$, d.h. eine Folge meßbarer Abbildungen $d_n : \mathcal{X}^n \longrightarrow \Gamma$

(5) eine Ellipsoidabbildung $e : \Gamma \longrightarrow \mathcal{E}(\mathcal{X})$, wobei

$$\mathcal{E}(\mathcal{X}) := \{T \in \mathcal{P}(\mathcal{X}) \mid \exists \text{ lineare bijektive Abbildung } A: \mathcal{X} \longrightarrow \mathcal{X} \\ \text{und } b \in \mathcal{X} \text{ mit } T = \{x \in \mathcal{X} \mid \|Ax + b\|^2 \leq 1\} \} \cup \{\mathcal{X}\}$$

die Menge der Ellipsoide in $\mathcal{X}$ ist.

(6) eine Projektionsabbildung

$$p : \mathcal{X} \times \mathcal{E}(\mathcal{X}) \longrightarrow \mathcal{X} \\ (x, E) \longmapsto p(x),$$

wobei

$$p(x) \begin{cases} = x, & x \in E \\ \in \partial E, & x \notin E \end{cases}$$

(7) eine meßbare Konsistenzkorrekturabbildung $k : \Gamma \longrightarrow \Gamma$.

Ist dann für $n \in \mathbb{N}$ und $(x_1, \dots, x_n) \in \mathcal{X}^n$ der Wert γ_0 eine Lösung von

$$d_n(p(x_1, e(\gamma_0)), \dots, p(x_n, e(\gamma_0))) = \gamma_0,$$

so heißt $k(\gamma_0)$ eine durch Winsorisieren mittels e , p und k aus d_n gewonnene Schätzung für $\gamma(P)$ auf der Basis der Stichprobe $(x_1, \dots, x_n)$. Ist $d_n^r : \mathcal{X}^n \longrightarrow \Gamma$ eine Folge meßbarer Abbildungen mit

$$d_n(p(x_1, e(d_n^r(x_1, \dots, x_n))), \dots, p(x_n, e(d_n^r(x_1, \dots, x_n)))) \\ = d_n^r(x_1, \dots, x_n)$$

für P^n -f.a. $(x_1, \dots, x_n) \in \mathcal{X}^n$ für jedes $P \in \mathcal{W}$, so heißt die Schätzerfolge $k \circ d_n$ eine durch Winsorisieren mittels e , p und k aus d_n gewonnene Schätzerfolge für $\gamma(P)$.

1.2.2. Spezialisierung:

zu (1): $(\mathcal{X}, \mathcal{B})$ wird hier stets $(\mathbb{R}^p, \mathcal{L}^p)$ sein. Als Norm werden

die Maximumnorm $\|\cdot\|_\infty$

und die Euklidische Norm $\|\cdot\|_2$

betrachtet. (Für $p \geq 2$ und $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}^p, \mathcal{L}^p)$ kann $\|\cdot\|_\infty$

nicht durch ein Skalarprodukt erhalten werden.)

zu (2): In der Regel wird $\mathcal{W}$ eine Lokations-Skalen-Familie sein.

zu (3): $(\Gamma, \mathcal{G})$ wird stets ein Teilraum von $(\mathbb{R}^\ell, \mathcal{L}^\ell)$ sein und i.d.R.

den Parameterraum Θ einer parametrischen Lokations-Skalen-Familie $(P_\vartheta)_{\vartheta \in \Theta}$ darstellen. Dann ist $\gamma(P_\vartheta) = \vartheta$ für alle P_ϑ aus der Familie $(P_\vartheta)_{\vartheta \in \Theta}$.

zu (4): Daß hier von der Robustifizierung gegebener - statt der direkten Definition neuer robuster - Schätzer ausgegangen wird, trägt der Vielfalt von Schätzern für manche Zwecke Rechnung (z.B. für Varianzkomponenten: ANOVA-Schätzer, ML-Schätzer, Hartung-Schätzer, etc.). Aber auch für den Fall der eindimensionalen Varianz gibt es etwa die unterschiedlichen Schätzer $\frac{1}{n} \sum (x_i - \bar{x})^2$ und $\frac{1}{n-1} \sum (x_i - \bar{x})^2$.

zu (5): Ein Ellipsoid in $(\mathbb{R}^p, \mathcal{L}^p)$ wird meist durch eine

$$\text{Mittelwertfunktion } m : \Gamma \longrightarrow \mathbb{R}^p$$

und eine

$$\text{Scatterfunktion } S : \Gamma \longrightarrow \text{NND}(p) \text{ (S p.d. } P^{\frac{n}{n-1}} \text{ f.s.)}$$

in der Form $E = \{x \mid \|S^{-1}(x - m)\|^2 \leq c\}$ beschrieben. Dabei liegen m und S meist kanonisch nahe (z.B. im wesentlichen als Identitäten im Falle der direkten Schätzung von Lokation und Scatter). Die Konstante c fungiert als "tuning constant". Hierfür wird i.a. der Wert eingesetzt werden, der E zu einem sogenannten "Konzentrationsellipsoid" macht. Grundsätzlich könnte man stets $c = 1$ durch entsprechende Wahl von S erreichen; aus Gründen der Interpretationsfähigkeit erscheint aber die obige Beschreibung praktischer.

zu (6): Als Projektion auf ein Ellipsoid E wird insbesondere die radiale Projektion auf den Mittelpunkt hin

$$p_r(x) = \frac{1}{\|S^{-1}(x-m)\|} \cdot (x - m) + m \quad \text{für } x \notin E$$

oder die Projektion in der Euklidischen Norm auf E

$$p_e(x) = x_0 \in E \quad \text{mit } \|x - x_0\|_2 = \text{Min!} \\ (\text{für festes } x \notin E)$$

gewählt.

zu (7): Die Konsistenzkorrektur betrifft i.a. nur den Skalierungsanteil von $\gamma(P)$. Sie dient dazu, die "Varianzverkleinerung" beim Winsorisieren auszugleichen und wird meist in der Multiplikation mit einer (von der Familie $(P_\vartheta)_{\vartheta \in \Theta}$ abhängigen) Konstanten bestehen.

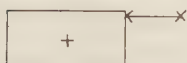
1.2.3. Bemerkungen:

- (1) Die radiale Projektion (vgl. "zu (6)") kann man auch als "Gewichtung" der Beobachtungen ansehen. Dafür könnte man noch sehr viel allgemeinere Gewichtungsfunktionen zulassen (z.B. redeszendierende). Ganz allgemein ließe sich dies durch eine "Beobachtungsveränderungsfunktion" $x_i \mapsto v(x_i)$ beschreiben. Da hier nur "klassisches" Winsorisieren verwendet wird, wird auf diese Verallgemeinerung kein Wert gelegt.
- (2) Die Wahl eines Konzentrationsellipsoides (vgl. "zu (5)") liegt nahe, da dies explizit zur Veranschaulichung des ersten und zweiten Momentes dienen soll (vgl. Cramér 1946, S. 300). Die Ellipsenform ist von der Norm abhängig, die "tuning Konstante" ist

$$c = p + 2 \quad (\text{für } \|\cdot\|_2)$$

$$c = 3 \quad (\text{für } \|\cdot\|_\infty).$$

- (3) Geometrisch ergeben sich die besonders interessanten Fälle



1) $\|\cdot\|_\infty, p_e$

2) $\|\cdot\|_\infty, p_r$

3) $\|\cdot\|_2, p_r$

Fall 1) kennzeichnet den "größten" und analytisch einfachsten, Fall 3) den "feinsten", aber analytisch schwierigsten Fall. Im eindimensionalen stimmen alle 3 Fälle überein.

- (4) Häufig wird d_n permutationsinvariant sein und - nach Identifizieren von x_i mit einer empirischen Verteilung (vgl. Hampel 1971) - als Einschränkung einer Funktion d auf einer geeigneten Menge $\mathcal{P}$ von W -Maßen auffaßbar sein. Wenn man dann für $P \in \mathcal{P}$ die auf $e(\gamma(P))$ komprimierte Verteilung (durch "Projektion" der Masse auf den Rand des Ellipsoides) erklärt, so kann man i.d.R. analog zum Vorgehen in 1.2.1. eine weitere Funktion auf $\mathcal{P}$ erklären. Bei gegebener Projektionsabbildung p , $\gamma \in \Gamma$, Ellipsoidabbildung e und $P \in \mathcal{P}$ schreiben wir dafür auch

$$d(p, e(\gamma), P).$$

Ist $d^\Gamma(p, e, \cdot): \mathcal{P} \rightarrow \Gamma$ eine Abbildung mit

$$d(p, e(d^\Gamma(p, e, P)), P) = d^\Gamma(p, e, P) \quad \forall P \in \mathcal{P},$$

so heißt $k \circ d^\Gamma(p, e, \cdot): \mathcal{P} \rightarrow \Gamma$ eine durch Winsorisieren

mittels e, p und k aus d gewonnene Funktion.

- (5) Durch die obigen Definitionen wird nichts über die Existenz, Eindeutigkeit und Robustheit von d_n^r oder d^r ausgesagt.
- (6) Rückgewinnungsmöglichkeit: Da $\mathcal{X} \in \mathcal{E}(\mathcal{X})$ erhält man durch $e: \Gamma \ni \gamma \mapsto \mathcal{X} \in \mathcal{E}(\mathcal{X})$ wieder den ursprünglichen Schätzer, d.h. $d_n^r = d_n$.

1.3. Allgemeine Aussagen über das Schätzprinzip

Ein wie in 1.2. robustifizierter Schätzer ergibt sich i.a. als Fixpunkt eines nichtlinearen Operators auf Γ . Daher lassen sich auch die Resultate der Fixpunkttheorie (vgl. z.B. Eisenack/Fenske (1978), Istrătescu (1981), Zeidler (1986)) nutzen. Im folgenden sind einige Resultate und Hilfsmittel zusammengestellt; Genaueres findet man jeweils bei den speziellen Anwendungen in Kapitel 2.-4.

1.3.1. Existenz:

Über die Existenz von Fixpunkten gibt es sehr umfangreiche Untersuchungen. Für unsere Zwecke ist i.a. der folgende Satz ausreichend:

1.3.1.1. Satz: Sei in der Situation aus 1.2. $\Gamma \subset \mathbb{R}^\ell$. Die Abbildung

$$\gamma \mapsto d_n(p(x_1, e(\gamma)), \dots, p(x_n, e(\gamma))) \quad (10)$$

sei stetig und bilde eine nichtleere kompakte konvexe Teilmenge $C \subset \Gamma \subset \mathbb{R}^\ell$ in sich ab. Dann existiert ein Fixpunkt von (10).

Beweis: Brouwerscher Fixpunktsatz.

1.3.2. Eindeutigkeit:

Die Eindeutigkeit des Fixpunktes (10) ist i.a. schwierig nachzuweisen, da meist der Banachsche Fixpunktsatz nicht anwendbar ist (s. 2.4.4.4. (ii)). Für gewisse W -Maße Q folgt sie aber aus Symmetriebetrachtungen. Setzt man dann mit den Bezeichnungen aus 1.2. in der Situation der Funktion (1.2.3. (4))

$$d(p, e(\gamma), t \cdot P + (1-t) \cdot Q) - \gamma =: H(\gamma, t), \quad (11)$$

wobei $P, Q \in \mathcal{P}$ und $\{tP + (1-t)Q \mid t \in [0, 1]\} \subset \mathcal{P}$, so kann man häufig die Eindeutigkeit durch eine Homotopieüberlegung erhalten. Wir

benutzen dazu die folgenden Überlegungen, die wir mit Blick auf spätere Anwendungen mit Mitteln der "nonsmooth Analysis" formulieren. Dabei folgen wir F.H. Clarke (1983, S. 70ff).

1.3.2.1. Situation: Eine Funktion $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ erfülle in einer Umgebung $\mathcal{U}(x)$ einer Stelle $x \in \mathbb{R}^n$ eine Lipschitz-Bedingung. Dann läßt sich zeigen ("Rademacher's Theorem"), daß eine Menge Ω_F vom Lebesgue-Maß $\lambda^n(\Omega_F) = 0$ existiert derart, daß F auf $\mathcal{U}(x) \setminus \Omega_F$ differenzierbar ist. (Einen Beweis oder Beleg gibt Clarke nicht, s. z.B. Federer 1969, Theorem 3.1.6., S. 216). Mit $JF(y)$ sei die $(m \times n)$ -Jacobi-Matrix von F für eine Stelle $y \in \mathcal{U}(x) \setminus \Omega_F$ bezeichnet.

1.3.2.2. Definition: Die verallgemeinerte Jacobi-Matrix von F an der Stelle x - in Zeichen: $\partial F(x)$ - ist die konvexe Hülle aller $(m \times n)$ -Matrixen, die sich als Grenzwert einer Folge der Form $JF(x_i)$ mit $x_i \rightarrow x$ und $x_i \in \Omega_F$ darstellen lassen. Wir sagen, $\partial F(x)$ hat Höchststrang (bzw. ist positiv definit (p.d.) bzw. negativ definit (n.d.)), wenn jedes Element von $\partial F(x)$ Höchststrang hat (bzw. $m = n$ und die hierdurch gegebene quadratische Form p.d. bzw. n.d. ist).

1.3.2.3. Bemerkung: In der obigen Definition hängt $\partial F(x)$ auch noch von Ω_F ab. Im Fall $m > 1$ ist unbekannt, ob unterschiedliche Festlegungen von Ω_F zu verschiedenen $\partial F(x)$ führen (s. F.H. Clarke 1983, S. 71). Clarke (1983, S. 71) beweist jedoch, daß der Rang von $\partial F(x)$ invariant bleibt. (Dies folgt aus Proposition 2.6.4.) Da es im folgenden stets nur auf den Rang ankommt, kann ohne Einschränkung angenommen werden, daß Ω_F in jeder Situation fest gewählt ist. (In vielen Fällen wird Ω_F auch "kanonisch" festliegen.)

1.3.2.4. Satz (Eigenschaften von ∂F): $B_{m \times n}$ sei die offene Einheitskugel im Matrizenraum $\mathbb{R}^{m \times n}$, der mit der Norm

$$\|M\|_{m \times n} = \left[\sum_{i=1}^m |r_i|^2 \right]^{1/2}$$

r_i ist die i -te Zeile von M und
 $|\cdot|$ die Euklidische Norm

versehen sei. Dann gilt:

- (a) $\partial F(x)$ ist eine nichtleere konvexe kompakte Teilmenge von $\mathbb{R}^{m \times n}$.
- (b) ∂F ist abgeschlossen an der Stelle x (d.h. falls $x_i \rightarrow x$, $Z_i \in \partial F(x_i)$ und $Z_i \rightarrow Z$, so folgt $Z \in \partial F(x)$).

(c) ∂F ist von oben halbstetig bei x (d.h. $\forall \varepsilon > 0 \exists \delta > 0 \forall y$ mit $|y-x| < \delta : \partial F(y) \subset \partial F(x) + \varepsilon B_{\max}$).

Beweis: F.H. Clarke (1983, Proposition 2.6.2., S. 70).

1.3.2.5. Satz (Existenz und Eindeutigkeit via Homotopie): Es sei Γ eine offene Teilmenge des $\mathbb{R}^\ell$, $H : \Gamma \times [0,1] \rightarrow \mathbb{R}^\ell$ eine Abbildung und $L := \{(\gamma', t') \in \Gamma \times [0,1] \mid H(\gamma', t') = 0\}$. Ferner gelte:

- (i) H ist stetig.
- (ii) $\exists$ kompakte Menge $M \subset \Gamma$ derart, daß für jedes $(\gamma', t') \in L$ gilt: $\gamma' \in M$.
- (iii) Zu jedem $(\gamma', t') \in L$ gibt es eine Umgebung $\mathcal{U}(\gamma', t')$, auf der H eine Lipschitz-Bedingung erfüllt.
- (iv) Für jedes $(\gamma', t') \in L$ hat $\pi_{\gamma'} \partial H(\gamma', t')$ Höchststrang. (Dabei bezeichne $\pi_{\gamma'} \partial H(\gamma', t')$ die Menge aller $(\ell \times \ell)$ -Matrizen M , für die es eine $(\ell \times 1)$ -Matrix N gibt, so daß die $(\ell \times (\ell + 1))$ -Matrix $[M, N]$ ein Element von $\partial H(\gamma', t')$ ist, vgl. F.H. Clarke 1983, S. 256.)

Besitzt dann die Gleichung $H(\gamma, 0) = 0$ genau eine Lösung $\gamma_0 \in \Gamma$, so besitzt auch die Gleichung $H(\gamma, 1) = 0$ genau eine Lösung $\gamma_1 \in \Gamma$.

Beweis: Mit einem Standardargument zeigt man zunächst, daß man von jeder Lösung $\gamma_0 =: g(0)$ der Gleichung $H(\gamma, 0) = 0$ auf einem Lipschitz-stetigen Pfad $g : [0,1] \rightarrow \Gamma$ zu einer Lösung $\gamma_1 =: g(1)$ der Gleichung $H(\gamma, 1) = 0$ gelangen kann. Sei dazu

$$A := \{t \in [0,1] \mid \exists \text{ Lipschitz-Funktion } g : [0,t] \rightarrow \Gamma \text{ mit} \\ H(g(u), u) = 0 \quad \forall u \in [0,t]\}$$

Dann gilt:

- (a) A ist nicht leer, da $0 \in A$.
- (b) A ist (relativ) offen in $[0,1]$ (d.h. offen bzgl. der Relativtopologie). Ist nämlich $t_1 \in A$ und $t_1 < 1$, so läßt sich ein Pfad $g_1 : [0,t] \rightarrow \Gamma$ nach dem verallgemeinerten Impliziten Funktionen Theorem (s. F.H. Clarke 1983, S. 256) auch noch auf eine Umgebung von t_1 fortsetzen.
- (c) A ist (relativ) abgeschlossen in $[0,1]$. Ist nämlich $(t_n)_{n \in \mathbb{N}}$ eine Folge mit $t_n \in A \quad \forall n \in \mathbb{N}$ und (o.B.d.A.) $t_n \uparrow t$, so existiert eine Folge von Lipschitz-Pfaden $g_n : [0, t_n] \rightarrow \Gamma$ mit $H(g_n(t_n), t_n) = 0$. Wegen (ii) existiert eine Teilfolge $(t'_n)_{n \in \mathbb{N}}$ der Folge (t_n) , so daß die zugehörige Teilfolge $g'_n(t'_n)$ gegen einen Wert $g(t) \in \Gamma$ konvergiert, für den wegen (i)

folgt: $H(g(t), t) = 0$. Nach (iv) und dem verallgemeinerten Impliziten Funktionen Theorem werden in einer geeignet gewählten Umgebung von $(g(t), t)$ alle Nullstellen von H durch einen Lipschitz-Pfad, etwa $f_2: [t-\varepsilon, t+\varepsilon] \cap [0, 1] \rightarrow \Gamma$, beschrieben. [Aus F.H. Clarkes (1983, S. 256) Formulierung geht die Eindeutigkeitsaussage des Impliziten Funktionen Theorems nicht direkt hervor; sie folgt aber aus dem Beweis, der wie im klassischen Fall mittels des Umkehrsatzes erfolgt.] Wählt man nun n_0 so, daß $t'_n > t - \frac{\varepsilon}{2} \forall n \geq n_0$ und einen Lipschitz-Pfad $f_1: [0, t'_n] \rightarrow \Gamma$ mit $H(f_1(u), u) = 0 \forall u \in [0, t'_n]$, so ergibt sich durch Aneinanderfügen von f_1 und f_2 , daß $t \in A$. Weil das Intervall $[0, 1]$ zusammenhängend ist, folgt aus (a) - (c), daß $A = [0, 1]$.

Die Existenz von γ_1 folgt somit direkt aus der Existenz von γ_0 . Gäbe es umgekehrt zwei Lösungen $(\gamma'_1, 1)$ und $(\gamma''_1, 1)$, so kann man durch Fortsetzung entlang der Lipschitz-Pfade g'_1 und g''_1 hieraus zwei Lösungen $(\gamma'_0, 0)$ und $(\gamma''_0, 0)$ gewinnen. Nach Voraussetzung ist aber $\gamma'_0 = \gamma''_0 = \gamma_0$. Daher muß der Punkt $(g'_1(t_0), t_0) = (g''_1(t_0), t_0)$ mit

$$t_0 := \inf \{t \in [0, 1] \mid g'_1(t) \neq g''_1(t)\}$$

ein Bifurkationspunkt sein. Dies widerspricht (iii) und (iv).

1.3.2.6. Korollar: In Satz 1.3.2.5. können die Voraussetzungen (iii) und (iv) durch die folgenden Bedingungen ersetzt werden:

(iii') Zu jedem $(\gamma', t') \in L$ gibt es eine Umgebung $U(\gamma', t')$, auf der H_γ als partielle Fréchet-Ableitung existiert und stetig ist.

(iv') Für jedes $(\gamma', t') \in L$ ist $H_\gamma(\gamma', t')$ bijektiv.

1.3.2.7. Notiz: Das Korollar bleibt auch gültig, wenn man Γ und $\mathbb{R}^l$ durch allgemeine Banachräume ersetzt. Für statistische Anwendungen wird dies jedoch nicht benötigt (Γ = Parameterraum); hingegen kann man die Verallgemeinerung auf den "nonsmooth" Fall gelegentlich zur Übertragung von Resultaten vom "stetigen" auf den "diskreten" Fall nutzen.

1.3.3. Berechenbarkeit:

Grundsätzlich lassen sich hierfür wieder alle Fixpunkt- oder Null-

stellenverfahren in Betracht ziehen. In praktischen Fällen kommt man aber bereits mit einfachen Verfahren (z.B. "einfache Iteration") aus oder benötigt "maßgeschneiderte" Methoden. Deswegen soll an dieser Stelle nur auf die prinzipielle numerische Behandelbarkeit eingegangen werden. Dazu bieten sich (vgl. 1.3.2.5.) Kontinuitätsverfahren (Pfadverfolgungsmethoden) an.

1.3.3.1. Algorithmus: Ausgehend von einer Lösung γ_0 von $H(\gamma_0, 0) = 0$ soll durch Verfolgen eines Pfades $\gamma = \gamma(t)$, $t \in [0, 1]$, eine Lösung γ_1 von $H(\gamma_1, 1) = 0$ gewonnen werden. Dazu sei $0 = t_0 < t_1 < \dots < t_N = 1$ eine Zerlegung von $[0, 1]$ und man betrachte die Folge von Newton-Iterationen:

$$\left. \begin{aligned} \gamma^{i, k+1} &= \gamma^{i, k} - \left[H_{\gamma}(\gamma^{i, k}, t_i) \right]^{-1} \cdot H(\gamma^{i, k}, t_i) & k = 0, \dots, m_i - 1 \\ \gamma^{1, 0} &= \gamma_0, \quad \gamma^{i+1, 0} = \gamma^{i, m_i} & i = 1, \dots, N-1 \\ \gamma^{N, k+1} &= \gamma^{N, k} - \left[H_{\gamma}(\gamma^{N, k}, 1) \right]^{-1} \cdot H(\gamma^{N, k}, 1) & k = 0, 1, \dots \end{aligned} \right\} \quad (12)$$

1.3.3.2. Satz: Die Abbildung $H : \Gamma \times [0, 1] \rightarrow \mathbb{R}^{\ell}$ ($\Gamma \subset \mathbb{R}^{\ell}$, offen) sei Fréchet-differenzierbar und die allgemeine partielle Ableitung nach γ H_{γ} sei auf $\Gamma \times [0, 1]$ stetig. Ferner seien die Voraussetzungen (ii) aus 1.3.2.5. und (iv') aus 1.3.2.6. erfüllt. Dann gibt es eine Zerlegung $0 = t_0 < \dots < t_N = 1$ von $[0, 1]$ und ganze Zahlen $m_1, \dots, m_{N-1}$, so daß die durch (12) definierte Folge stets in Γ bleibt und $\lim_{k \rightarrow \infty} \gamma^{N, k} = \gamma_1$ ist.

Beweis: Ortega/Rheinboldt (1970, 10.4.2., S. 336). (Die Existenz eines stetigen Pfades $\gamma : [0, 1] \rightarrow \Gamma$ mit $H(\gamma(t), t) = 0 \quad \forall t \in [0, 1]$ folgt wie im Beweis von 1.3.2.5. (mit dem "klassischen" Impliziten Funktionen Theorem).)

1.3.3.3. Bemerkung:

- (i) Satz 1.3.3.2. macht nur eine prinzipielle Aussage, da die Zerlegung und die Zahlen m_j nicht konstruiert werden.
- (ii) Es ist möglich, die Aussage des Satzes auf den Stichprobenfall auszuweiten, wenn man fordert, daß die im Beweis benutzten Aussagen über die Jacobi-Matrizen jeweils für alle Matrizen aus der verallgemeinerten Jacobi-Matrix im Sinne von 1.3.2.2. gelten. Dafür gibt es aber keine griffigen Kri-

terien (wie z.B. Stetigkeit von H_γ). Ein Ausweg besteht in vielen Fällen darin, den Stichprobenfall durch "Verschmieren" der Beobachtungen auf einer Kreisscheibe mit Radius ε_n stetig zu machen. Betrachtet man eine Folge $\varepsilon_n \downarrow 0$, so kann man jeweils "prinzipiell" die einzelnen Probleme wie oben lösen und erhält schließlich als Grenzwert die gesuchte Lösung. Die Konvergenz wird sich häufig aus der Struktur des Problems (Verteilungskonvergenz) ergeben. Allerdings kann man das Konvergenzproblem vernachlässigen, wenn man eine endliche Rechengenauigkeit (z.B. Maschinengenauigkeit ε) unterstellt. Dann unterscheiden sich die Probleme von einem hinreichend großen n an nicht mehr.

- (iii) Der Algorithmus (12) ist Spezialfall einer sogenannten "Predictor-Corrector" Methode (s. z.B. Garcia/Zangwill 1981, S. 310ff), die sich aus einer Prädiktionsstufe (Iteration über i) und einer Korrekturstufe (Iteration über k) zusammensetzen. Im Falle des Algorithmus (12) handelt es sich um "horizontales" Korrigieren, da t_1 bei der Iteration über k fest bleibt. Grenzfall solcher Verfahren ist auch die oben erwähnte "einfache Iteration" (nur "Korrekturstufe").

Manchmal ist folgender Sonderfall von Interesse:

1.3.3.4. Bemerkung: Es seien die Voraussetzungen von Satz 1.3.2.5. erfüllt. Der dann existierende Lipschitz-stetige Pfad $\gamma: [0,1] \rightarrow \Gamma$ habe die Lipschitz-Konstante K . Ferner gebe es eine Iterationsvorschrift $it: \Gamma \rightarrow \Gamma$ mit der Eigenschaft:

$$\exists \varepsilon > 0 \quad \forall t \in [0,1] \quad \forall \gamma^* \text{ mit } \|\gamma^* - \gamma(t)\| \leq \varepsilon \quad \exists n \in \mathbb{N}: \gamma(t) = it^n(\gamma^*), \quad (13)$$

wobei it^n die n -malige Hintereinanderführung von it bezeichnet. Dann läßt sich die exakte Lösung $\gamma_1 := \gamma(1)$ in endlich vielen Schritten berechnen.

Beweis: Setze $N := \lceil \frac{K}{\varepsilon} \rceil + 1$, $t_0 := 0$, $t_{i+1} := t_i + \frac{1}{N}$ für $i = 0, \dots, N-1$. Für jedes $i = 0, \dots, N-1$ gilt dann: Ist $\gamma(t_i)$ bekannt, so ist $\gamma^* := \gamma(t_i)$ ein geeigneter Startwert für die Korrekturiteration an der Stelle t_{i+1} , denn es ist $\|\gamma(t_i) - \gamma(t_{i+1})\| < K \cdot |t_i - t_{i+1}| \leq \varepsilon$. Somit kann man in endlich vielen Schritten $\gamma(t_{i+1}) = it^n(\gamma(t_i))$ berechnen. Da das Prädiktorinkrement fest und positiv ist und γ_0 bekannt ist, folgt die Be-

hauptung.

1.3.4. Suffizienz:

Da Suffizienz und Robustheit sehr verschiedene Konzepte sind (vgl. z.B. Lehmann 1983, S. 36) ist d_n^r i.a. nicht Funktion suffizienter Statistiken T. Wir notieren nur den einfachen

1.3.4.1. Spezialfall: Es sei $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}, \mathcal{L})$, und $x^{(n)} = (x_1, \dots, x_n)$ seien Realisationen von n unabhängigen Zufallsvariablen X_i mit identischer Verteilung $P \in \mathcal{P}$. Ferner sei $\pi^n: \mathbb{R}^n \rightarrow \mathbb{R}^n$ eine Permutation der x_i und es gelte $d_n(\pi^n(x^{(n)})) = d_n(x^{(n)})$. Dann läßt sich d_n^r als Funktion der suffizienten Statistik $\pi^n(x^{(n)})$ schreiben. Insbesondere gilt dies für die Ordnungsstatistik $x_{[1]}^{(n)}$ (vgl. Witting 1985, S. 336f).

1.3.5. Equivarianz:

Zum Hintergrund von Equivarianzüberlegungen sei auf Lehmann (1983, S. 165ff) verwiesen.

1.3.5.1. Situation: $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ sei eine identifizierbare Familie (d.h. $P_\theta = P_{\theta'} \Rightarrow \theta = \theta'$) von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$. Ferner sei G eine Gruppe von bijektiven bimeßbaren Abbildungen: $\mathcal{X} \rightarrow \mathcal{X}$, die $\mathcal{P}$ invariant läßt. Die zugehörige Gruppe von Abbildungen $\bar{g}: \Theta \rightarrow \Theta$ sei mit $\bar{G}$ bezeichnet. Schließlich bezeichne G^* die bei der Schätzung von $\gamma(P) =: \gamma(\theta) \in \Gamma$ auftretende Gruppe von Abbildungen $g^*: \Gamma \rightarrow \Gamma$, wobei (11) und (14) aus Lehmann (1983, S. 167) erfüllt seien.

1.3.5.2. Satz: Es liege die in 1.3.5.1. beschriebene Situation vor. Ferner gelte

$$\forall \gamma \in \Gamma \quad \forall g \in G \quad \forall x \in \mathcal{X} : p(g(x), e(g^*(\gamma))) = g(p(x, e(\gamma))). \quad (14)$$

(i) Ist dann d_n ein equivarianter Schätzer unter $g \in G$ (d.h. es ist $d_n(g(x_1), \dots, g(x_n)) = g^*(d_n(x_1, \dots, x_n)) \quad \forall (x_1, \dots, x_n) \in \mathcal{X}^n$), so gilt die Äquivalenz:

$$\begin{aligned} d_n(p(x_1, e(\gamma)), \dots, p(x_n, e(\gamma))) &= \gamma \\ \Leftrightarrow d_n(p(g(x_1), e(g^*(\gamma))), \dots, p(g(x_n), e(g^*(\gamma)))) &= g^*(\gamma). \end{aligned}$$

(ii) Ist die Existenz und Eindeutigkeit von d_n^r gesichert, so ist

mit d_n auch d_n^r equivariant unter $g \in G$.

Beweis: (i) $d_n(p(x_1, e(\gamma)), \dots, p(x_n, e(\gamma))) = \gamma$

$$\begin{aligned} &\Leftrightarrow d_n(p(g(x_1), e(g^*(\gamma))), \dots, p(g(x_n), e(g^*(\gamma)))) \\ &= d_n(g(p(x_1, e(\gamma))), \dots, g(p(x_n, e(\gamma)))) \\ (14) \quad &= g^*(d(p(x_1, e(\gamma)), \dots, p(x_n, e(\gamma)))) \\ &= g^*(\gamma) \end{aligned}$$

(ii) folgt aus (i).

1.3.5.3. Notiz: Die in 1.3.5.1. skizzierte Situation gibt den statistischen Hintergrund wieder. Der Satz benötigt nur die Existenz von Abbildungsgruppen G und G^* mit den entsprechenden Eigenschaften. Ähnliche Aussagen für die Funktion aus 1.2.3.(4) könnte man angeben, wenn man die Gruppen $\bar{G}$ auf Θ (betrachtet als Gruppe auf $\mathcal{P}$) und G^* betrachtet.

1.3.6. Statistische Eigenschaften:

Wir stellen einige Hilfsmittel bereit, die in konkreten Situationen zum Nachweis statistischer Eigenschaften dienen können:

1.3.6.1. Konsistenz: Hierzu dienen die Bedingungen A und B aus Huber (1981, S. 128-132).

1.3.6.2. Asymptotische Normalität: Hierzu dienen die Bedingungen aus Huber (1981, S. 132f). Insbesondere Corollary 3.2 kann zur Berechnung der asymptotischen Kovarianzmatrix benutzt werden.

1.3.6.3. Stetigkeit - Differenzierbarkeit - Robustheit: Hierzu gibt es einige Untersuchungen über M-Schätzer, z.B. diejenigen von Fernholz (1983), und über von Mises-Funktionale, z.B. diejenigen von Boos/Serfling (1980) und B.R. Clarke (1986).

1.3.7. Klassenbildung:

Durch die Schätzung wird die Menge $\mathcal{P}$ in Klassen zerlegt.

1.3.7.1. Definition: Ist $d^r(p, e, \cdot)$ eine Funktion auf $\mathcal{P}$ wie in 1.2.3.(4), so heißt für $\gamma \in \Gamma$

$$[d^r(p, e, \cdot)]^{-1}(\{\gamma\}) := \{P \in \mathcal{P} \mid d^r(p, e, P) = \gamma\}$$

die zu γ gehörige Klasse in $\mathcal{P}$, in Zeichen: $Kl(\gamma; \mathcal{P})$. Ist $\mathcal{P}$ in einer Situation unmißverständlich festgelegt, so schreiben wir auch kurz $Kl(\gamma)$.

1.3.7.2. Bemerkung:

- (i) Die Relation $P \sim Q :\Leftrightarrow d^r(p, e, P) = d^r(p, e, Q)$ ist eine Äquivalenzrelation in $\mathcal{P}$. Statistisch sinnvoll ist diese Relation vor allem, falls $d^r(p, e, \cdot)$ (als Funktion) "kanonisch" eindeutig bestimmt ist.
- (ii) Equivarianzüberlegungen (vgl. 1.3.5.) gestatten es oft, sich auf die Untersuchung einer speziellen Klasse $Kl(\gamma)$ zu beschränken.

Wir verallgemeinern noch den Begriff der Maximalmasse (vgl. Krein (1959, S. 33) oder Krein/Nudel'man (1977, S. 70ff) oder Karlin/Studden (1966, S. 49f)):

1.3.7.3. Definition: Es sei $d^r(p, e, \cdot)$ eine Funktion auf $\mathcal{P}$ wie in 1.2.3.(4) und $Kl(\gamma; \mathcal{P}) \neq \emptyset$ die zu einem $\gamma \in \Gamma$ gehörige Klasse. Dann heißt die Abbildung $M : \mathcal{X} \rightarrow [0, 1]$

$$x \mapsto \sup_{Q \in Kl(\gamma; \mathcal{P})} Q(\{x\})$$

die Maximalmassenfunktion von $Kl(\gamma; \mathcal{P})$. Ferner heißt

$$M_u := \inf_{x \in \mathcal{X}} M(x) \quad [\text{bzw. } M_o := \sup_{x \in \mathcal{X}} M(x)]$$

die untere [bzw. obere] Maximalmasse von $Kl(\gamma; \mathcal{P})$ und

$$M_B := M_o - M_u$$

die Maximalmassenbandbreite von $Kl(\gamma; \mathcal{P})$. Wir sagen, daß M "beschränkten Träger besitzt", falls es eine beschränkte Menge $T \subset \mathcal{X}$ gibt mit

$$M(x) = M^T(x) := \sup_{\substack{Q \in Kl(\gamma; \mathcal{P}) \\ \text{supp } Q \subset \{x\} \cup T}} Q(\{x\}).$$

1.3.7.4. Bemerkung: Die Maximalmassenfunktion veranschaulicht die Auswirkungen der Robustifizierung; sie hat insofern eine ähnliche Rolle wie die Influenzfunktion (vgl. 2.5.2.2. für eine genauere

Illustration). $M(x_0)$ gibt die größte Masse an, die auf x_0 gelegt werden kann, ohne die jeweilige Klasse $Kl(\gamma; \mathcal{P})$ zu verlassen, und entspricht daher dem intuitiven Konzept des "breakdown point". Speziell M_u ist mit dem "gross-error breakdown point" ϵ (Hampel e.a. 1986, S. 97) verwandt; für die in den Abschnitten 2.2., 2.3. und 2.4. betrachteten Fälle läßt sich jeweils leicht verifizieren:

- (i) $\epsilon \leq M_u$
- (ii) $\epsilon = M_u$, falls M beschränkten Träger besitzt.

Falls M beschränkten Träger besitzt, bezeichnen wir daher M_u als "Breakdown-Punkt". Der "Breakdown-Punkt" ist also i.w. nichtprobabilitisch definiert, was aber durchaus mit den Vorläuferideen (Hodges 1967, S. 180ff) oder modernen Betrachtungen (Donoho/Huber 1983, S. 159ff) in Einklang steht.

1.3.8. Hilfsmittel zur Darstellung:

Die Klassen $Kl(\gamma; \mathcal{P})$ aus 1.3.7.1. sind i.d.R. konvexe Mengen, die durch Momentenbedingungen spezialisiert sind. Zu ihrer Darstellung kann man oft die folgenden Hilfsmittel nutzen:

1.3.8.1. Momentenproblem: Es diesbezüglich auf die Bücher von Shohat/Tamarkin (1943), Karlin/Studden (1966) und Krein/Nudel'man (1977) verwiesen.

1.3.8.2. Extrempunktdarstellungen (nach Choquet): Es wird vor allem Theorem 2 (bzw. 7) aus Mulholland/Rogers (1958, S. 178) benutzt. Eine wesentlich allgemeinere Fassung geben etwa Weizsäcker/Winkler (1980, Proposition 5, S. 138).

2. LOKATIONS- UND SKALENSCHÄTZUNG IM EINDIMENSIONALEN FALL

Ziel dieses Kapitels ist nicht eine genaue Erörterung des robusten Schätzens von Lokation und Skalenwert im eindimensionalen Fall, vgl. hierfür die ausführlichen Darlegungen in Huber (1981, Ch. 3-6), Hampel e.a. (1986, Ch. 2) oder Lehmann (1983, Ch. 5). Vielmehr soll die Anwendung des Prinzips aus Kapitel 1 in einem übersichtlichen Fall dargestellt und mit bekannten Resultaten verglichen werden.

2.1. Schätzung eines eindimensionalen Parameters

Im folgenden sei γ ein eindimensionaler Parameter und Γ ein offenes Intervall in $\mathbb{R}$. Dann vereinfachen sich viele Untersuchungen, wenn man Monotonieüberlegungen heranziehen kann.

2.1.1. Definition: Es sei $d(p, e(\gamma), \cdot)$ ein Funktional auf $\mathcal{P}$ wie in 1.2.3.(4). Ist $P \in \mathcal{P}$ und $\gamma_0 \in \Gamma$ eine Stelle mit $d(p, e(\gamma_0), P) = \gamma_0$, so heißt $d(p, e(\cdot), P)$ für die Verteilung P im Fixpunkt γ_0 fallend [bzw. steigend], falls es ein offenes Intervall $]\gamma_0 - \varepsilon, \gamma_0 + \varepsilon[\subset \Gamma$ gibt mit der Eigenschaft

$$\forall \gamma \in]\gamma_0 - \varepsilon, \gamma_0[: d(p, e(\gamma), P) > \gamma \quad [\text{bzw.} < \gamma]$$

$$\forall \gamma \in]\gamma_0, \gamma_0 + \varepsilon[: d(p, e(\gamma), P) < \gamma \quad [\text{bzw.} > \gamma]$$

2.1.2. Bemerkung: Es sei $d(p, e(\gamma), \cdot)$ ein Funktional auf $\mathcal{P}$ wie in 1.2.3.(4), und es gelte für alle $P \in \mathcal{P}$:

- (i) $d(p, e(\cdot), P) : \Gamma \rightarrow \Gamma$ ist stetig
- (ii) $\exists \gamma_1, \gamma_2 \in \Gamma$ mit $d(p, e(\gamma_1), P) - \gamma_1 \leq 0 \leq d(p, e(\gamma_2), P) - \gamma_2$
- (iii) $d(p, e(\cdot), P)$ ist in jedem Fixpunkt fallend [bzw. in jedem Fixpunkt steigend].

Dann gibt es für jedes $P \in \mathcal{P}$ genau einen Fixpunkt $\gamma(P) \in \Gamma$ von

$d(p, e(\cdot), P)$, und das Funktional $d^r(p, e, \cdot)$ auf $\mathcal{P}$ ist eindeutig bestimmt.

2.1.3. Bemerkung: Hinreichend für (iii) in 2.1.2. ist

(iii') $F(\gamma) := d(p, e(\gamma), P) - \gamma$ ist in der Nähe eines jeden Fixpunktes γ_0 von $d(p, e(\cdot), P)$ Lipschitz-stetig, und es ist dort jeweils $\partial F(\gamma_0) = \partial\{d(p, e(\gamma_0), P) - \gamma_0\}$ n.d. [bzw. jeweils p.d.].

2.1.4. Satz: Es seien die Voraussetzungen aus 2.1.2. erfüllt. Ferner gelte für ein festes $P \in \mathcal{P}$ und jedes feste $\gamma \in \Gamma$

$$d(p, e(\gamma), P^n) \rightarrow d(p, e(\gamma), P) \quad \text{f.s.}, \quad (1)$$

wobei P^n die empirische Verteilung von n i.i.d. Zufallsvariablen mit der Verteilung P bezeichne. Dann ist $d^r(p, e, P^n)$ stark konsistent für $d^r(p, e, P)$.

Beweis: Es sei o.B.d.A. $d(p, e(\cdot), P)$ in jedem Fixpunkt fallend. Nach 2.1.2. gibt es ein eindeutig bestimmtes γ_0 , so daß für hinreichend kleines $\varepsilon > 0$ gilt

$$d(p, e(\gamma_0 - \varepsilon), P^n) \rightarrow d(p, e(\gamma_0 - \varepsilon), P) > \gamma_0 - \varepsilon \quad \text{f.s.}$$

$$d(p, e(\gamma_0 + \varepsilon), P^n) \rightarrow d(p, e(\gamma_0 + \varepsilon), P) < \gamma_0 + \varepsilon \quad \text{f.s.}$$

Es folgt $\gamma_0 - \varepsilon \leq \liminf_{n \rightarrow \infty} d^r(p, e, P^n) \leq \limsup_{n \rightarrow \infty} d^r(p, e, P^n) \leq \gamma_0 + \varepsilon$ f.s. und durch den Grenzübergang $\varepsilon \rightarrow 0$ die Behauptung.

2.1.5. Situation: Zur Untersuchung der asymptotischen Eigenschaften von Schätzerfolgen hat von Mises (1947) die Differentialrechnung auf statistische Funktionale angewendet. Dabei ergeben sich allerdings einige technische Schwierigkeiten - insbesondere bei der Wahl des Differenzierbarkeitsbegriffs. Diesbezüglich sei auf Serfling (1980, Ch. 6), und Fernholz (1983) verwiesen. Um Anschluß an die Resultate von Fernholz (1983) zu gewinnen, benutzen wir folgende Notation: $C[0,1]$ ist der Raum der stetigen reellwertigen Funktionen auf $[0,1]$ versehen mit der uniformen Topologie (d.i. die von der Supremumsnorm $\|G\| := \sup_{x \in [0,1]} |G(x)|$ (für $G \in C[0,1]$) induzierte Topologie). Ferner sei $D[0,1]$ der Raum der rechtsseitig stetigen reellwertigen Funktionen auf $[0,1]$, die an jeder Stelle einen linksseitigen Limes besitzen, ebenfalls versehen mit der uniformen Topologie. F

sei eine feste Verteilungsfunktion und T ein für F erklärtes statistisches Funktional, d.h. eine reellwertige Abbildung auf einer konvexen Menge von Verteilungsfunktionen, die mindestens F und die empirischen Verteilungsfunktionen F^n enthält. Dann wird durch die Definition

$$\tau(G) := T(G \circ F) \quad (\text{für } G \in C[0,1] \text{ bzw. } D[0,1])$$

ein Funktional auf einer Umgebung der Verteilungsfunktion der Gleichverteilung in $C[0,1]$ bzw. $D[0,1]$ definiert. In etwas nachlässiger Sprechweise heißt τ das von T auf $C[0,1]$ bzw. $D[0,1]$ induzierte Funktional. Es ist notationsmäßig bequem, i.f. nicht streng zwischen Verteilungsfunktionen und den zugehörigen Verteilungen zu unterscheiden.

2.1.6. Satz: Es sei $\mathcal{P}$ Obermenge einer konvexen Menge von Verteilungsfunktionen auf $(\mathbb{R}, \mathcal{B})$, die wenigstens eine stetige, streng monoton steigende Verteilungsfunktion F und deren empirische Verteilungsfunktionen F^n enthält, und es seien die Voraussetzungen aus 2.1.2. erfüllt. Ferner sei $d^r(p, e, \cdot)$ das nach 2.1.2. eindeutig bestimmte Funktional auf $\mathcal{P}$ und τ das in 2.1.5. erklärte induzierte Funktional auf [einer Teilmenge von] $D[0,1]$. Ist dann τ auf einer offenen Umgebung der Gleichverteilung U erklärt und an der Stelle U Hadamard-differenzierbar mit der Ableitung τ'_u , wobei gilt

$$0 < \sigma^2 := \text{Var}_F(\tau'_u((\delta_x - F) \circ F^{-1})) < \infty, \quad (2)$$

so gilt:

$$\sqrt{n} (T(F^n) - T(F)) \xrightarrow{\mathcal{D}} N(0, \sigma^2). \quad (3)$$

Zusatz: Es ist $\tau'_u((\delta_x - F) \circ F^{-1}) = IC(x; F, T)$. Dabei bezeichnet δ_x das Dirac-Maß auf x und IC die Influenzkurve.

Beweis: Fernholz (1983, Theorem 4.4.2., S. 40 und Lemma 4.4.1., S. 39).

2.1.7. Bemerkung:

- (i) Ein ähnliches Resultat existiert auch für das auf $C[0,1]$ induzierte Funktional (s. Fernholz 1983, Corollary 4.4.3., S. 41).
- (ii) I.a. wird $d^r(p, e, P)$ als Lösung t einer Gleichung der Form

$$\int \psi(x, t) dP(x) = 0 \quad (4)$$

vorliegen (als sogenannter "M-Schätzer"). Für diesen spezielleren Fall gibt es weitere Untersuchungen, in denen z.T. auch LIL- oder Berry-Esséen-Schranken abgeleitet oder Entwicklungen untersucht werden, vgl. insbesondere Serfling (1980), B.R. Clarke (1986), Carroll (1978a), Stute (1986).

2.2. Schätzung eines Lokationsparameters

2.2.1. Situation: Es liegt eine Lokationsfamilie $F_\mu(x) = F(x - \mu)$ vor, und es soll μ (robust) geschätzt werden. (Der Skalenwert σ ist bekannt und habe o.E. den Wert 1.) Die Projektionen und Ellipsen aus 1.2.3.(3) stimmen dann überein; es werden i.f. stets $\|\cdot\|_2$ und p_r benutzt (o.B.d.A.) und in der Notation weitgehend unterdrückt. Ein Ellipsoid um einen Lokationswert t_0 ist stets ein Intervall $I(t_0)$ mit $t_0 \in I(t_0)$.

2.2.2. Robustifizierung des Mittelwertes:

Der kanonische Schätzer für μ ist $d_n = \frac{1}{n} \sum x_i$ bzw. als Funktional $d(F) = \int x dF(x)$. Ein kanonisches Ellipsoid ist $I(t_0) = [t_0 - k, t_0 + k]$ (Mittelwertfunktion = Identität, Scatterfunktion = 1, "tuning constant" = k^2). Damit ergibt sich der Huber-Schätzer aus 1.1. Die Wahl von k ist grundsätzlich offen und hängt vom gewünschten Grad der Robustifizierung ab (vgl. Huber 1964, S. 84, Table I) Die Grenzfälle sind $k = \infty$ (Mittel) und $k = 0$ (Median). Falls F_0 bekannt ist, könnte man etwa k so wählen, daß $F_0([-k, k]) = q$, wobei q eine fest gewählte Konstante ist (z.B. $q = 0.95$), oder so, daß $\int_{[-k, k]} x^2 dF_0 + k^2 \cdot (1 - F_0([-k, k])) = \frac{k^2}{3}$ erfüllt ist. Ist F_0 unbekannt, erscheint die Wahl des "Konzentrationsellipsoides" $I(t_0) = [t_0 - \sqrt{3}, t_0 + \sqrt{3}]$ aus datenanalytischer Sicht nicht unplausibel (dies entspricht $F_0 = U$, vgl. auch 2.5. unten). Mit den Notationen aus 1.1. und 2.1. ist zunächst $\Gamma = \mathbb{R}$ und für ein W-Maß P auf $(\mathbb{R}, \mathcal{L})$ gilt

$$d(p, e(t), P) =: d(t, P) = t + \int_{\mathbb{R}} \psi(x-t) dP(x) = t - Q'(t).$$

Die Bedingungen (i) und (ii) aus 2.1.2. sind trivialerweise erfüllt (s. 1.1.5.(iii)) und (iii') aus 2.1.3. besagt gerade, daß die Funktion Q' die reelle Achse im Fixpunkt t_0 mit positiver

"Steigung" durchstößt. Zur Charakterisierung derjenigen W-Maße, die (iii') erfüllen, sei zunächst bemerkt, daß mit ψ (als Funktion von t) auch Q' auf $\mathbb{R}$ Lipschitz-stetig ist (für jedes W-Maß P). Zerlegt man P in die Maße

$$R_1 := r_1 \cdot \delta_{t_0 - k} + r_3 \cdot \delta_{t_0 + k}$$

$$R_2 := P - R_1$$

mit $r_1 := P(\{t_0 - k\})$, $r_3 := P(\{t_0 + k\})$, $\delta_x :=$ Dirac-Maß auf x , so ergibt sich mit $r_2 := R_2([t_0 - k, t_0 + k]) = P([t_0 - k, t_0 + k])$

$$\partial\{d(t_0, P) - t_0\} = \text{co}\{-r_2 - r_3, -r_2 - r_1\},$$

wobei $\text{co } M$ die konvexe Hülle der Menge M sei. Ein hinreichendes Kriterium für die Eindeutigkeit ist also

$$\text{Max}\{-r_2 - r_3, -r_2 - r_1\} < 0$$

bzw. abgeschwächt:

$$P([t_0 - k, t_0 + k]) > 0 \text{ für jeden Fixpunkt } t_0 \text{ von } d(p, e(\cdot), P). \quad (5)$$

Durch dieses Kriterium werden nur diejenigen Maße P zu Unrecht nicht erfaßt, für die es eine Stelle t_0 gibt mit:

$$\left. \begin{aligned} \text{(i)} \quad & P([-\infty, t_0 - k]) = \frac{1}{2} = P([t_0 + k, \infty]) \\ \text{(ii)} \quad & P([t_0 - k, t_0 + k]) = 0 \\ \text{(iii)} \quad & \forall \varepsilon > 0 : P([t_0 - k - \varepsilon, t_0 + k + \varepsilon]) > 0. \end{aligned} \right\} \quad (6)$$

Dies zeigt, daß nur "Grenzfälle" nicht abgedeckt werden (vgl. 1.1.7.(i)). Es sei

$$\mathcal{P}_0 := \left\{ \text{W-Maße } P \text{ auf } (\mathbb{R}, \mathcal{L}) \mid \forall t \in \mathbb{R}: P([-\infty, t - k]) \neq \frac{1}{2} \vee P([t - k, t + k]) > 0 \vee P([t + k, \infty]) \neq \frac{1}{2} \right\}.$$

Dann ist auf $\mathcal{P}_0$ das robustifizierte Funktional $d^r(\cdot)$ erklärt. Wegen des starken Gesetzes der großen Zahlen gilt für jedes $P \in \mathcal{P}_0$ und jedes feste t :

$$d(t, P^n) = \frac{1}{n} \sum \psi(x_i - t) \longrightarrow \int \psi(x - t) dP(x) = d(t, P) \quad P\text{-f.s.},$$

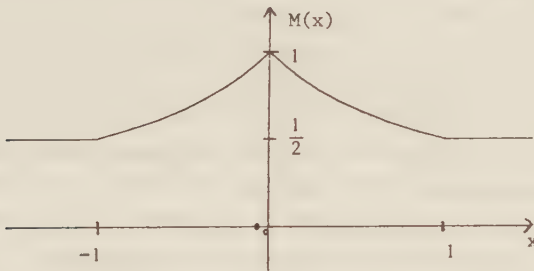
so daß nach 2.1.4. $d^r(P^n)$ stark konsistent für $d^r(P)$ ist.

Entsprechend folgt aus 2.1.6. die asymptotische Normalität (s. Fernholz 1983, Theorem 7.1.2., S. 91 und Example 7.1.3., S. 92), die man aber in diesem Fall einfacher direkt beweisen kann (s.

z.B. Huber 1981, Theorem 2.4., S. 50). Man beachte, daß die asymptotische Normalität nicht für alle $P \in \mathcal{P}_0$ gilt (Gegenbeispiel: δ_0). Für weitere Resultate vgl. die unter 2.1.7. zitierte Literatur. Schließlich ist $d^r(\cdot)$ translationsequivariant (equivariant unter der Gruppe der Translationen, vgl. 1.3.5.) und legt in $\mathcal{P}_0$ eine Äquivalenzrelation nach 1.3.7. fest. Im folgenden wird noch $K1(0) = K1(0; \mathcal{P}_0)$ genauer betrachtet, wobei o.B.d.A. $k=1$ gesetzt sei. Die Maximalmassenfunktion ist gegeben durch

$$M(x) = \begin{cases} \frac{1}{1+|x|} & , \quad |x| < 1 \\ \frac{1}{2} & , \quad |x| \geq 1 \end{cases}$$

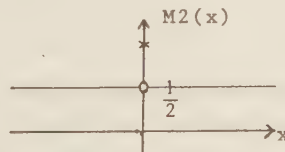
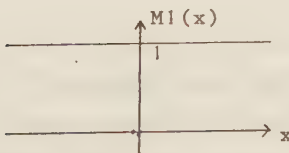
und läßt sich so skizzieren:



Dabei ist $M_0 = 1$, $M_u = \frac{1}{2}$, $M_B = \frac{1}{2}$. Offensichtlich ist M_u der höchste erreichbare Breakdown-Punkt für einen Lokationsschätzer (s. auch Huber 1981, S. 54). Zum Vergleich ist die Maximalmassenfunktion des Mittelwertes $M1$ bzw. des Medians $M2$ (jeweils für $K1(0)$ bezüglich der Menge $\mathcal{P}_1$ aller W -Maße auf $(\mathbb{R}, \mathcal{L})$ mit endlichem ersten Moment bzw. eindeutigem Median):

$$M1(x) \equiv 1$$

$$M2(x) = \begin{cases} 1, & x = 0 \\ \frac{1}{2}, & x \neq 0 \end{cases}$$



Dies spiegelt die in 1.1.7.(ii) festgestellte Interpretation des Huber-Schätzers als "korrigierter Median" wider. [Die scheinbar idealen Werte $M1_u = 1$ und $M1_B = 0$ für das Mittel deuten hier nicht

auf günstige Eigenschaften hin, weil M_1 nicht "beschränkten Träger" im Sinne von 1.3.7. besitzt. Würde man dort $T:=[-1,1]$ setzen, ergäbe sich $M^T(x) = \frac{1}{1+|x|}$, während die Maximalmassenfunktion von $d^r(\cdot)$ und dem Median unverändert bleiben. Dann ist $M_1^u = 0$ und $M_2^B = 1$. Als Momentenproblem betrachtet, ergibt sich (vgl. 1.3.8.1.):

- für das Mittel:

$$Kl(0) = \{P \in \mathcal{P}_1 \mid P \text{ löst das Hamburger-Momentenproblem mit } \mu_0 = 1 \text{ und } \mu_1 = 0\}$$

- für $d^r(\cdot)$:

$$Kl(0) = \{P \in \mathcal{P}_0 \mid \text{Die Einschränkung von } P \text{ auf } [-1,1] \text{ ist eine Lösung des Stieltjes-Momentenproblems auf } [-1,1] \text{ mit } \mu_0 = 1 - p_1 - p_3 \text{ und } \mu_1 = p_1 - p_3, \text{ wobei } p_1 := P((-\infty, -1]) \text{ und } p_3 := P((1, \infty))\}.$$

Die "Kennzahlen" p_1 und p_3 beschreiben die Klasse $Kl(0)$; ihr Laufbereich ist $[0, \frac{1}{2}] \times [0, \frac{1}{2}]$. Der Mittelwert μ_1 über $[-1,1]$ von P liegt dann in $]-\frac{1}{2}, \frac{1}{2}[$, während er für ein beliebiges substochastisches Maß den Bereich $[-1,1]$ abdeckt.

Schließlich kann man die konvexe Menge $Kl(0)$ für $d^r(\cdot)$ durch ihre Extrempunkte (nach 1.3.8.2.) beschreiben. Diese sind:

- das Dirac-Maß δ_0
- die Zwei-Punkt-Maße $P = p\delta_{x_1} + (1-p)\delta_{x_2}$ mit $p \neq 0$, $\psi(x_1) \neq \psi(x_2)$ und $\int \psi(x) dP(x) = 0$.

Um die Disjunktheit zweier Klassen $Kl(t)$ und $Kl(u)$ zu illustrieren, nimmt man zunächst o.B.d.A. $u = -t < 0$ an und wendet auf die Klassen das konvexe Funktional

$$a: P \longrightarrow \int \psi(x) dP(x)$$

an. Man erhält mit $m := \min(1, t)$:

$$a(Kl(t)) =]0, m]$$

$$a(Kl(-t)) = [-m, 0[$$

so daß die Klassen durch die "separierende Hyperebene" $\{0\}$ getrennt werden. Aus diesem Grund müssen bei der Definition von P_0 Maße wie in 1.1.7.(i) ausgeschlossen werden. Sie bilden gewissermaßen den gemeinsamen "Rand" der Klassen.

2.3. Schätzung eines Skalenparameters

2.3.1. Situation: Es liege eine Skalenfamilie $F_{\sigma}(x) = F(\frac{x}{\sigma})$ vor, und es soll σ (robust) geschätzt werden. (Der Lokationswert μ sei bekannt und habe o.E. den Wert 0.) Für die Projektionen und Ellipsen gilt das in 2.2.1. Gesagte.

2.3.2. Robustifizierung der Varianz:

Der kanonische Schätzer für σ ist die empirische Standardabweichung, die ihrerseits die Wurzel aus der empirischen Varianz

$$v_n = s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2$$

ist. (Da $\mu = 0$ bekannt ist, ist dies im Falle der Normalverteilung zugleich der ML-Schätzer.) Ein kanonisches Ellipsoid ist $I(v) = [-k\sqrt{v}, k\sqrt{v}]$ (Mittelwertfunktion $\equiv 0$, Scatterfunktion = Quadratwurzel, "tuning constant" = k^2), wobei wie in 2.2.2. $k = \sqrt{3}$ für das Konzentrationsellipsoid gewählt werden muß. (In analoger Weise könnte man durch einen Parameterwechsel das Problem auch im Parameter $s := \sqrt{v}$ betrachten, dann ist die Scatterfunktion die Identität.) Mit den Notationen aus 1.1. und 2.1. ist dann $\Gamma =]0, \infty[$ und für ein W-Maß P auf $(\mathbb{R}, \mathcal{L})$ gilt

$$d(p, e(v), P) =: d(v, P) = \int x^2 dP_{0, k\sqrt{v}}(x) = v \cdot \int [\psi(\frac{x}{\sqrt{v}})]^2 dP(x).$$

Dabei bezeichne $P_{t,c}$ hier und im folgenden das auf $[t - c, t + c]$ konzentrierte Maß, das analog wie in 1.1.7.(iii) definiert sei. Die Bedingungen (i) und (ii) aus 2.1.2. sind offensichtlich genau dann gleichzeitig erfüllt, falls

$$\exists a > 0 : P([-a, a]) \leq \frac{k^2 - 1}{k^2}. \quad (7)$$

Zum Nachweis von (iii') aus 2.1.3. erhält man in analoger Weise wie in 2.2.

$$\begin{aligned} \partial\{d(v_0, P) - v_0\} &= \text{co}\{k^2 - 1 - k^2 \cdot P([-k\sqrt{v_0}, k\sqrt{v_0}]), \\ &\quad k^2 - 1 - k^2 \cdot P([-k\sqrt{v_0}, k\sqrt{v_0}])\} \end{aligned}$$

Eine hinreichende Bedingung für die Eindeutigkeit ist also

$$P([-k\sqrt{v_0}, k\sqrt{v_0}]) > \frac{k^2 - 1}{k^2} \quad (8)$$

für jeden Fixpunkt v_0 . Andererseits gilt für einen Fixpunkt v_0 stets

$$P(\{]-k\sqrt{v_0}, k\sqrt{v_0}[\}) \geq \frac{k^2-1}{k^2}, \quad (9)$$

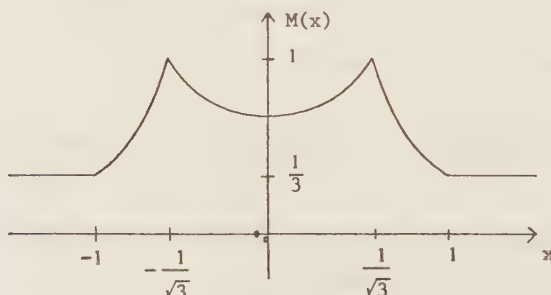
da sonst $\int x^2 dP_{0, k\sqrt{v_0}}(x) > \frac{1}{k^2} \cdot k^2 \cdot v_0 = v_0$ mit Widerspruch zur Fixpunkteigenschaft folgte. Durch (8) werden zur Sicherung der Eindeutigkeit also nur diejenigen W-Maße P ausgeschlossen, für die es ein $v_0 > 0$ gibt mit $P(\{]-k\sqrt{v_0}, k\sqrt{v_0}[\}) = \frac{k^2-1}{k^2} = P(\{0\})$. Für derartige Maße gilt in der Tat die Eindeutigkeit nicht. (Jedes $v < v_0$ ist auch Fixpunkt.) Zusammen mit der für die Existenz erforderlichen Bedingung (7) ist also auf der Menge

$$\mathcal{P}_0 = \left\{ \text{W-Maße auf } (\mathbb{R}, \mathcal{L}) \mid P(\{0\}) < \frac{k^2-1}{k^2} \right\} \quad (10)$$

das robustifizierte Funktional $d^r(\cdot)$ in eindeutiger Weise erklärt. Aussagen zur Konsistenz und asymptotischen Normalität folgen wie in 2.2.2. $d^r(\cdot)$ ist ersichtlich skalenequivariant und legt in $\mathcal{P}_0$ eine Äquivalenzrelation nach 1.3.7. fest. Im folgenden wird für die Wahl $k := \sqrt{3}$ noch die Klasse $Kl(\frac{1}{3}) = Kl(\frac{1}{3}; \mathcal{P}_0)$ genauer beschrieben (d.h. es ist $v_0 = \frac{1}{3}$ und damit $\sqrt{3} \cdot \sqrt{v_0} = 1$). Die Maximalmassenfunktion ist gegeben durch

$$M(x) = \begin{cases} \frac{2}{3(1-x^2)} & , \quad |x| < \frac{1}{\sqrt{3}} \\ \frac{1}{3x^2} & , \quad \frac{1}{\sqrt{3}} \leq |x| \leq 1 \\ \frac{1}{3} & , \quad 1 < |x| \end{cases}$$

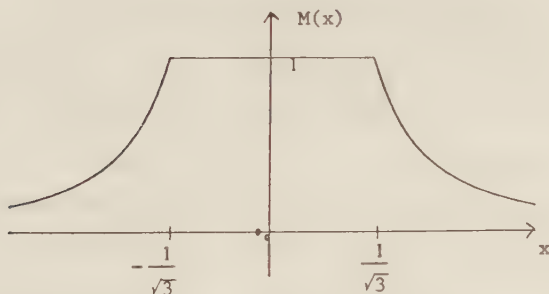
und läßt sich wie folgt skizzieren:



Es ist $M_0 = 1$, $M_u = \frac{1}{3}$, $M_B = \frac{2}{3}$. Der Breakdown-Punkt ist offenbar $M_u = \frac{1}{3}$, da M etwa den beschränkten Träger $[-1, 1]$ besitzt. Die

Maximalmassenfunktion der Varianz ist für $Kl(\frac{1}{3})$ (in der Menge $\mathcal{P}_2$ aller W-Maße auf $(\mathbb{R}, \mathcal{L})$ mit endlichem zweiten Moment) zum Vergleich

$$M(x) = \begin{cases} 1 & , |x| \leq \frac{1}{\sqrt{3}} \\ \frac{1}{3x^2} & , |x| > \frac{1}{\sqrt{3}} \end{cases}$$



Es ist dann $M_0 = 1$, $M_u = 0$, $M_B = 1$. Auch diese Funktion hat den beschränkten Träger $[-1, 1]$ und somit den Breakdown-Punkt 0. Als Momentenproblem betrachtet, ergibt sich (vgl. 1.3.8.1.)

- für die Varianz:

$$Kl(\frac{1}{3}) = \left\{ P \in \mathcal{P}_2 \mid P \text{ löst das Hamburger-Momentenproblem mit } \mu_0 = 1, \mu_2 = \frac{1}{3} \right\}$$

- für $d^r(\cdot)$:

$$Kl(\frac{1}{3}) = \left\{ P \in \mathcal{P}_0 \mid \begin{array}{l} \text{Die Einschränkung von } P \text{ auf } [-1, 1] \text{ ist eine} \\ \text{Lösung des Stieltjes-Momentenproblems auf } [-1, 1] \\ \text{mit } \mu_0 = 1 - p_1 - p_3 \text{ und } \mu_2 = \mu_0 - \frac{2}{3}, \text{ wobei } p_1 = \\ P(-\infty, -1]) \text{ und } p_3 = P(1, \infty) \end{array} \right\}.$$

Die Kennzahlen p_1 und p_3 durchlaufen den Bereich $\{(p_1, p_3) \mid p_1 + p_3 < \frac{1}{3}\}$. Das zweite Moment μ_2 über $[-1, 1]$ von P liegt

dann in $[0, \frac{1}{3}]$, während für ein beliebiges substochastisches Maß der Bereich $[0, 1]$ abgedeckt wird. Schließlich kann man die konvexe Menge $Kl(\frac{1}{3})$ wieder nach 1.3.8.2. durch ihre Extrempunkte (Dirac-Maße auf $\pm \frac{1}{\sqrt{3}}$ und gewisse Zwei-Punkt-Maße) beschreiben.

2.4. Schätzung eines Lokations- und Skalenparameters

2.4.1 Situation: Es liege eine Lokations- und Skalenfamilie $F_{\mu, \sigma}(x) = F\left(\frac{x-\mu}{\sigma}\right)$ vor, und es sollen μ und σ (robust) geschätzt werden. Für die Projektionen und Ellipsen gilt wieder das in 2.2.1. Gesagte.

2.4.2. Robustifizierung von Mittelwert und Varianz:

Die kanonischen Schätzer für μ und σ^2 sind $m_n = \frac{1}{n} \sum x_i$ und $s_n^2 = \frac{1}{n-1} \sum (x_i - m_n)^2$. Im Sinne einer dann möglichen funktionalen Interpretation ist es jedoch (zunächst) zweckmäßiger, die Schätzer

$$m^u(P) = \int x dP(x) \quad \text{und} \quad v^u(P) = \int (x - m^u(P))^2 dP(x)$$

(u = "unrobustifiziert") zugrunde zu legen. Ein kanonisches Ellipsoid ist $I(\gamma) = I(m, v) = [m - k\sqrt{v}, m + k\sqrt{v}]$ (Mittelwertfunktion = Identität, Scatterfunktion = Quadratwurzel, "tuning constant" = k^2), wobei wieder $k = \sqrt{3}$ für das Konzentrationsellipsoid zu wählen ist. Dann ist $\gamma = (m, v) \in \Gamma =]-\infty, \infty[\times]0, \infty[$, und für ein W -Maß P auf $(R, \mathcal{E})$ gilt:

$$d(p, e(\gamma), P) =: d((m, v), P) = \begin{bmatrix} \int x dP_{m, k\sqrt{v}}(x) \\ \int (x - m)^2 dP_{m, k\sqrt{v}}(x) \end{bmatrix} = \begin{bmatrix} m + \int \psi(x - m) dP(x) \\ v \cdot \int \psi\left(\frac{x - m}{\sqrt{v}}\right)^2 dP(x) \end{bmatrix}. \quad (11)$$

Gesucht sind die Fixpunkte (m_0, v_0) der Abbildung

$$d : (m, v) \longrightarrow d((m, v), P) \quad (12)$$

aus (11). Es sei Q eine Verteilung mit einem eindeutig bestimmten Fixpunkt (m'_0, v'_0) . (Für eine um 0 symmetrische Verteilung muß z.B. aus Symmetriegründen $m'_0 = 0$ sein, und die Eindeutigkeit von v'_0 folgt aus 2.3.) Es sei W ein weiteres W -Maß auf $(R, \mathcal{E})$. Wir betrachten die Abbildung (11) aus 1.3.2.

$$H(\gamma, t) := d(p, e(\gamma), t \cdot W + (1-t) \cdot Q) - \gamma \quad (13)$$

Die Eigenschaften (i) und (iii) aus 1.3.2.5. sind erfüllt, da die Integranden an jeder Stelle Lipschitz-stetig sind. Setzt man daher zusätzlich die Gültigkeit von (ii) und (iv) voraus, so folgt die Existenz und Eindeutigkeit eines Fixpunktes von (12) für W aus 1.3.2.5. Die folgenden Sätze und Bemerkungen stellen hierzu Kriterien bereit.

2.4.2.1. Bemerkung (Reduktion durch Equivarianz): Es sei $((m_0, v_0), t_0) = (\gamma_0, t_0)$ eine beliebige feste Nullstelle von H aus (13), und es seien W' und Q' die durch

$$W'(B) := W \left\{ \left\{ \frac{x - m_0}{k\sqrt{v_0}} \mid x \in B \right\} \right\} \quad \forall B \in \mathcal{L}$$

$$Q'(B) := Q \left\{ \left\{ \frac{x - m_0}{k\sqrt{v_0}} \mid x \in B \right\} \right\} \quad \forall B \in \mathcal{L}$$

gegebenen W -Maße. Ferner sei $P := t_0 \cdot W + (1 - t_0) \cdot Q$ und $P' := t_0 \cdot W' + (1 - t_0) \cdot Q'$. Mit den Transformationen

$$T'(m, v) := \begin{bmatrix} \frac{m - m_0}{k\sqrt{v_0}} \\ \frac{v}{k^2 v_0} \end{bmatrix} \quad \text{und} \quad T''(m, v) := \begin{bmatrix} k\sqrt{v_0} \cdot m + m_0 \\ k^2 \cdot v_0 \cdot v \end{bmatrix}$$

gilt dann die Beziehung

$$d((m, v), P) = T''(d(T'(m, v), P')).$$

Es ist also $((0, \frac{1}{k^2}), t_0)$ eine Nullstelle von

$$H'(\gamma, t) := d(p, e(\gamma), t \cdot W' + (1 - t) \cdot Q') - \gamma,$$

und es gilt weiter

$$H(\gamma, t) = T''(H'(T'(\gamma), t)) - \begin{bmatrix} m_0 \\ 0 \end{bmatrix}$$

und

$$\pi_{\gamma}(\partial H(\gamma_0, t_0)) = \left\{ \begin{bmatrix} k\sqrt{v_0} & 0 \\ 0 & k^2 v_0 \end{bmatrix} \cdot D \cdot \begin{bmatrix} \frac{1}{k\sqrt{v_0}} & 0 \\ 0 & \frac{1}{k^2 v_0} \end{bmatrix} \mid D \in \pi_{\gamma}(\partial H'((0, \frac{1}{k^2}), t_0)) \right\}. \quad (14)$$

Aus diesem Grunde kann man sich zur Herleitung von Regularitätskriterien für $\pi_{\gamma}(\partial H(\gamma_0, t_0))$ o.B.d.A. auf den Fall $m_0 = 0$ und $v_0 = \frac{1}{k^2}$ beschränken. Dann ist $k\sqrt{v_0} = 1$.

2.4.2.2. Bemerkung (Momentenbedingungen):

(a) Eine notwendige und hinreichende Bedingung für die Existenz eines Maßes M auf $[-1, 1]$ mit den ersten drei (zentralen) Momenten (μ_0, μ_1, μ_2) ist:

$$\mu_0 \geq 0, \mu_2 \geq 0, \mu_0 \geq \mu_2, \mu_0 \cdot \mu_2 \geq \mu_1^2. \quad (15)$$

(b) Es sei $(\gamma_0, t_0) = ((0, \frac{1}{k^2}), t_0)$ eine Nullstelle von H aus (13).

Mit den Bezeichnungen

$$\begin{aligned} P &:= t_0 \cdot W + (1-t_0) \cdot Q \\ p_1 &:= P(-\infty, -1[) \\ r_1 &:= P(\{-1\}) \\ p_2 &:= P(-1, 1[) \\ r_3 &:= P(\{1\}) \\ p_3 &:= P(]1, \infty[) \end{aligned} \quad (16)$$

gilt dann:

$$(i) \quad k \geq 1$$

$$(ii) \quad p_1 + r_1 \leq \frac{1}{k^2+1}, \quad p_3 + r_3 \leq \frac{1}{k^2+1}$$

$$(iii) \quad \frac{k^2+1}{k^2} \geq 4 \cdot \frac{1}{k^2+1}.$$

Beweis: (a) Die Aussage folgt unmittelbar aus den bekannten Sätzen zum Momentenproblem (s. z.B. Shohat/Tamarkin 1943, S. 77, für die hier benötigte Form) und dem Hauptminorenkriterium für positive Semidefinitheit von Matrizen. (Letzteres wird allerdings gelegentlich falsch angegeben, z.B. in Beckenbach/Bellman 1965, S. 59; die korrekte Fassung gibt Gantmacher 1986, S. 316.)

(b) Es gibt in diesem Falle ein Maß auf $] -1, 1[$ und daher auch auf $[-1, 1]$ mit den ersten drei Momenten $(1-p_1-r_1, p_1+r_1, \frac{1}{k^2} - p_1-r_1)$. Unter Benutzung von (15) folgt hieraus (i) und der erste Teil von (ii). Der zweite Teil von (ii) folgt aus Symmetriegründen. (iii) ist klar wegen (i).

2.4.2.3. Satz: Es sei $(\gamma_0, t_0) = ((0, \frac{1}{k^2}), t_0)$ eine Nullstelle von H aus (13). Mit den Bezeichnungen (16) gilt dann:

(a) Es ist $\pi_\gamma(\partial H((0, \frac{1}{k^2}), t_0)) = \text{co}\{D_1, D_2, D_3, D_4\}$, wobei $D_i, i=1, \dots, 4$, wie unter (17), (18), (19) und (20) gegeben sind.

(b) Folgende Aussagen sind äquivalent:

(i) $(-1) \cdot \pi_\gamma(\partial H((0, \frac{1}{k^2}), t_0))$ ist regulär.

(ii) Für die ersten drei Momente $\mu_i := \int_{]-1, 1[} x^i dP(x) \quad (i=1, 2, 3)$

der Einschränkung von P auf $] -1, 1[$ gilt: (μ_0, μ_1, μ_2) ist innerer Punkt des Momentenraums $\mathcal{M}_3$ über $[-1, 1]$ (vgl. Karlin/Studden 1966, S. 106ff).

- (iii) Es gibt kein $a \in] -1, 1[$ mit $\text{supp}(P) \cap] -1, 1[= \{a\}$, d.h. das auf $] -1, 1[$ eingeschränkte Maß ist nicht zum Einpunktmaß entartet.

Beweis: (a) Durch Nachrechnen ergibt sich als

- Ableitung für $(m+k\sqrt{v})\uparrow 1$ und $(m-k\sqrt{v})\downarrow -1$:

$$D_1 := \begin{pmatrix} -p_2 & \frac{k^2}{2}(p_3 + r_3 - p_1 - r_1) \\ 2(p_3 + r_3 - p_1 - r_1) & k^2(\frac{k^2-1}{k^2} - p_2) \end{pmatrix} \quad (17)$$

- Ableitung für $(m+k\sqrt{v})\downarrow 1$ und $(m-k\sqrt{v})\downarrow -1$:

$$D_2 := \begin{pmatrix} -p_2 - r_3 & \frac{k^2}{2}(p_3 - p_1 - r_1) \\ 2(p_3 - p_1 - r_1) & k^2(\frac{k^2-1}{k^2} - p_2 - r_3) \end{pmatrix} \quad (18)$$

- Ableitung für $(m+k\sqrt{v})\uparrow 1$ und $(m-k\sqrt{v})\uparrow -1$:

$$D_3 := \begin{pmatrix} -p_2 - r_1 & \frac{k^2}{2}(p_3 + r_3 - p_1) \\ 2(p_3 + r_3 - p_1) & k^2(\frac{k^2-1}{k^2} - p_2 - r_1) \end{pmatrix} \quad (19)$$

- Ableitung für $(m+k\sqrt{v})\downarrow 1$ und $(m-k\sqrt{v})\uparrow -1$:

$$D_4 := \begin{pmatrix} -p_2 - r_3 - r_1 & \frac{k^2}{2}(p_3 - p_1) \\ 2(p_3 - p_1) & k^2(\frac{k^2-1}{k^2} - p_2 - r_3 - r_1) \end{pmatrix} . \quad (20)$$

Andere Folgen (m_n, v_n) mit $(m_n, v_n) \rightarrow (0, \frac{1}{k^2})$ führen zu keinem oder einem der oben angegebenen Grenzwerte.

(b) Zunächst ist

$$\mu_0 = p_2, \quad \mu_1 = p_1 + r_1 - p_3 - r_3, \quad \mu_2 = p_2 - \frac{k^2-1}{k^2} .$$

Setzt man

$$\tilde{D} := \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_0 & \mu_2 \end{pmatrix} = \begin{pmatrix} p_2 & p_1 + r_1 - p_3 - r_3 \\ p_1 + r_1 - p_3 - r_3 & p_2 - \frac{k^2 - 1}{k^2} \end{pmatrix}, \quad (21)$$

so gelten offensichtlich nach bekannten Sätzen über das Momentenproblem (s. z.B. Karlin/Shapley 1953, S. 22 und S. 56f) die Äquivalenzen

$$(iii) \Leftrightarrow (ii) \Leftrightarrow \tilde{D} \text{ ist p.d.}$$

Die Determinante einer Konvexkombination $K = \sum_{i=1}^4 \lambda_i D_i$ mit $\sum_{i=1}^4 \lambda_i = 1$ ist

$$\det K = k^2 \cdot \det \sum_{i=1}^4 \lambda_i \tilde{D}_i,$$

wobei für $i=1, \dots, 4$

$$\tilde{D}_i := \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \cdot D_i \cdot \begin{pmatrix} 2 & 0 \\ 0 & k^2 \end{pmatrix}^{-1}$$

sei. Weil $\tilde{D}$ sicherlich positiv semidefinit ist, reicht es zum Beweis von (b) nach der Ky Fan-Ungleichung (vgl. Marshall/Olkin 1979, S. 476) aus, die Äquivalenz

$$\tilde{D} \text{ ist p.d.} \Leftrightarrow (-\tilde{D}_1) \text{ ist p.d. für } i=1, \dots, 4.$$

zu zeigen. Da die ersten Hauptminore stets positiv sind und $\det(\tilde{D}) = \det(-\tilde{D}_1)$ ist, bleibt nur

$$\det(-\tilde{D}_1) \geq \det(\tilde{D}) \text{ für } i=2, 3, 4$$

zu zeigen. Hierzu gilt zunächst:

$$\begin{aligned} & \det(-\tilde{D}_2) \geq \det \tilde{D} \\ \Leftrightarrow & r_3 = 0 \vee \frac{k^2 + 1}{k^2} + r_3 \geq 4(p_1 + r_1) \\ \Leftrightarrow & \frac{k^2 + 1}{k^2} \geq 4(p_1 + r_1). \end{aligned}$$

Letzteres ist wegen 2.4.2.2.(b) erfüllt. Analog folgt $\det(-\tilde{D}_3) \geq \det \tilde{D}$. Schließlich gilt:

$$\begin{aligned} & \det(-\tilde{D}_4) \geq \det \tilde{D} \\ \Leftrightarrow & \frac{k^2 + 1}{k^2} (r_1 + r_3) \geq 4(r_1 p_3 + r_3 p_1 + r_1 r_3) \end{aligned}$$

$$\Leftarrow \left(\frac{k^2+1}{k^2} - \frac{4}{k^2+1} + 4r_3 \right) r_1 + \left(\frac{k^2+1}{k^2} - \frac{4}{k^2+1} + 4r_1 \right) r_3 \geq 4r_1 r_3$$

$$\Leftarrow 8r_1 r_3 \geq 4r_1 r_3$$

2.4.2.2. (b)

Damit ist (b) bewiesen.

Wir stellen noch ein weiteres Hilfsmittel aus der Theorie des Momentenproblems bereit. Da dies von unabhängigem Interesse ist, formulieren wir den Hilfssatz ganz allgemein in der seit Krein (1959) üblichen Terminologie, wie sie z.B. auch in Karlin/Studden (1966) und Krein/Nudel'man (1977) verwendet wird.

2.4.2.4. Hilfssatz: Es sei $\{c_k\}_0^n$ eine strikt positive reelle Zahlenfolge, $[a,b] \subset \mathbb{R}$ ein abgeschlossenes Intervall und $V(c) = V(c_0, c_1, \dots, c_n)$ die Menge der Massenverteilungen $\sigma(t)$ auf $[a,b]$ mit $\int t^k d\sigma(t) = c_k$ für $k=0,1,\dots,n$. Dann enthält $V(c)$ wenigstens ein atomloses Maß.

Beweis: Es sei zunächst $n=2\nu-1$ eine ungerade Zahl. Wir benutzen die Ergebnisse von Krein/Nudel'man (1977, S.88ff) und parametrisieren den dort beschriebenen Prozeß des "Durchpumpens" von Masse mit Hilfe eines Laufparameters t über dem Intervall $[-\rho(a), \rho(b)]$. Für $t \in [-\rho(a), 0]$ (bzw. $t \in]0, \rho(b)]$) sei $H_t(x)$ die Verteilungsfunktion der kanonischen Darstellung σ_{μ_a} mit $\mu_a = |t|$ (bzw. σ_{μ_b} mit $\mu_b = t$). Dann ist für jedes feste t $H_t(\cdot)$ als Funktion von x eine Treppenfunktion mit $\nu+1$ Sprungstellen. Diese Sprungstellen hängen stetig von t ab. Die Sprunghöhen hängen daher auch stetig von t ab. Andererseits ist für jedes feste $x \in [a,b]$ $H_t(x)$ als Funktion von t stetig mit Ausnahme höchstens einer Sprungstelle. Somit gilt nach bekannten Sätzen über parameterabhängige Integrale (s. z.B. Hinderer 1972, Satz 19.9, S. 120), daß die Funktion

$$F(x) := \frac{1}{\rho(a)+\rho(b)} \int_{-\rho(a)}^{\rho(b)} H_t(x) dt$$

auf $[a,b]$ stetig ist. F ist somit die Verteilungsfunktion eines atomlosen Maßes aus $V(c)$.

Für den Fall einer geraden Zahl $n=2\nu$ verläuft der Beweis analog.

Trotz der engen Beziehung zum sogenannten $(0,L)$ -Momentenproblem

(vgl. Krein/Nudel'man 1977, Ch. VII, A 2) scheint ein Beweis von Hilfssatz 2.4.2.4. mit Hilfe des Grenzübergangs $L \rightarrow \infty$ nicht einfacher zu sein. Für den weiter unten benötigten Spezialfall läßt sich auch ein atomloses Maß explizit konstruieren:

2.4.2.5. Spezialfall: Es sei (μ_0, μ_1, μ_2) ein Tripel reeller Zahlen mit der Eigenschaft $\mu_0 > \mu_2 > 0$ und $\mu_0 \cdot \mu_2 > \mu_1^2$. Dann existiert ein atomloses Maß M auf $[-1, 1]$ mit den ersten drei zentralen Momenten (μ_0, μ_1, μ_2) .

Beweis: Wir können o.B.d.A. $\mu_0 = 1$ und $\mu_1 \geq 0$ annehmen. Für reelle Zahlen $p_1, p_2, p_3 \in [0, 1]$, $\ell \in]0, 2[$ und $k \in]0, 1 - \mu_1[$ ist dann durch die Lebesgue-Dichte

$$f(x) = \frac{p_1}{\ell} \cdot 1_{[-1, -1+\ell]}(x) + \frac{p_2}{2k} \cdot 1_{[\mu_1-k, \mu_1+k]}(x) + \frac{p_3}{\ell} \cdot 1_{[1-\ell, 1]}(x)$$

ein atomloses Maß auf $[-1, 1]$ definiert, das die geforderten Momente genau dann besitzt, wenn das Gleichungssystem

$$p_1 + p_2 + p_3 = 1$$

$$(p_3 - p_1)\left(1 - \frac{\ell}{2}\right) + p_2\mu_1 = \mu_1$$

$$(p_3 + p_1)(3 - 3\ell + \ell^2) + p_2(3\mu_1^2 + k^2) = 3\mu_2$$

erfüllt ist. Mit den neuen Variablen $x := p_3 - p_1$ und $y := p_3 + p_1$ ergibt sich das äquivalente System

$$p_2 = 1 - y$$

$$x\left(1 - \frac{\ell}{2}\right) = y\mu_1$$

$$y(3 - 3\ell + \ell^2 - 3\mu_1^2 - k^2) = 3(\mu_2 - \mu_1^2) - k^2,$$

und es ist hinreichend, ein $\ell \in]0, 2(1 - \mu_1)[$ und ein $k \in]0, 1 - \mu_1[$ mit der Eigenschaft

$$(3 - 3\ell + \ell^2 - 3\mu_1^2 - k^2) > 3(\mu_2 - \mu_1^2) - k^2 > 0$$

anzugeben, da sich dann geeignete Werte für y , x und p_2 sukzessive

aus den Gleichungen ergeben. Die Werte $k_0 := \left[\sqrt{\mu_2 - \mu_1^2} \right] (1 - \mu_1)$

und $\ell_0 := (1 - \mu_1)(1 - \mu_2)$ leisten das Verlangte.

2.4.2.6. Satz (Schachtelung): Es sei P ein W -Maß auf $(\mathbb{R}, \mathcal{B})$, und (m_1, v_1) sowie (m_2, v_2) mit $v_2 \leq v_1$ seien Fixpunkte von (12). Dann gilt:

$$[m_2 - k\sqrt{v_2}, m_2 + k\sqrt{v_2}] \subset [m_1 - k\sqrt{v_1}, m_1 + k\sqrt{v_1}].$$

Beweis: Wir können o.B.d.A. $m_1 < m_2$ annehmen (sonst spiegele man P um m_2), so daß $m_2 - k\sqrt{v_2} > m_1 - k\sqrt{v_1}$. Ferner gilt offensichtlich $k > 1$, $P([m_1 + k\sqrt{v_1}, \infty[) < \frac{1}{2}$ und $P(]-\infty, m_2 - k\sqrt{v_2}]) < \frac{1}{2}$. Es folgt daher $m_2 - k\sqrt{v_2} < m_1 + k\sqrt{v_1}$ und $P([m_2 - k\sqrt{v_2}, m_1 + k\sqrt{v_1}]) > 0$. Falls $m_2 + k\sqrt{v_2} \leq m_1 + k\sqrt{v_1}$, ist nichts zu zeigen. Falls $m_2 + k\sqrt{v_2} > m_1 + k\sqrt{v_1}$, so würde folgen:

$$\begin{aligned} 0 &= \int x \, dP_{m_1, k\sqrt{v_1}}(x) - m_1 - \int x \, dP_{m_2, k\sqrt{v_2}}(x) + m_2 \\ &\geq (m_2 - m_1) + P(]-\infty, m_2 - k\sqrt{v_2}[) \cdot (m_1 - k\sqrt{v_1} - m_2 + k\sqrt{v_2}) \\ &\quad + P([m_1 + k\sqrt{v_1}, \infty[) \cdot (m_1 + k\sqrt{v_1} - m_2 - k\sqrt{v_2}) \\ &= (m_2 - m_1) \cdot P([m_2 - k\sqrt{v_2}, m_1 + k\sqrt{v_1}]) + (k\sqrt{v_1} - k\sqrt{v_2}) \cdot \\ &\quad (P([m_1 + k\sqrt{v_1}, \infty[) - P(]-\infty, m_2 - k\sqrt{v_2}[)) \\ &\geq (m_2 - m_1) \cdot \min\{P([m_2 - k\sqrt{v_2}, \infty[) - P(]-\infty, m_2 - k\sqrt{v_2}[), \\ &\quad P(]-\infty, m_1 + k\sqrt{v_1}[) - P([m_1 + k\sqrt{v_1}, \infty[))\} \\ &\quad (\text{denn: } m_2 + k\sqrt{v_2} > m_1 + k\sqrt{v_1} \Rightarrow m_2 - m_1 > k\sqrt{v_1} - k\sqrt{v_2} \text{ (} \geq 0)) \\ &> 0, \end{aligned}$$

also ein Widerspruch.

2.4.2.7. Definition: Es sei P ein W -Maß auf $(\mathbb{R}, \mathcal{B})$ und (m, v) ein Fixpunkt von (12). Ist dann $\text{supp } P \cap]m - k\sqrt{v}, m + k\sqrt{v}] \neq \{a\}$ für alle $a \in \mathbb{R}$, so heißt (m, v) ein regulärer, andernfalls ein singulärer Fixpunkt von P .

2.4.2.8. Satz: Es sei P ein W -Maß auf $(\mathbb{R}, \mathcal{B})$, Q ein weiteres W -Maß auf $(\mathbb{R}, \mathcal{B})$ mit genau einem Fixpunkt (m', v') , und es gelte:

(a) Es gibt eine kompakte Menge $M \subset \Gamma$ mit

$$M \supset \{(m, v) \in \Gamma \mid \exists t \in [0, 1] \text{ derart, daß } (m, v) \text{ Fixpunkt von } t \cdot P + (1-t) \cdot Q \text{ ist}\}.$$

(b) Für jedes $t \in [0, 1]$ besitzt das Maß $t \cdot P + (1-t) \cdot Q$ nur reguläre Fixpunkte.

Dann besitzt P genau einen regulären Fixpunkt.

Beweis: Für die Abbildung H aus (13) sind die Voraussetzungen des Satzes 1.3.2.5. zu verifizieren. (i) und (iii) folgen aus der Lipschitz-Stetigkeit der Integranden in (11). (iii) ist gleichbedeutend mit (a); (iv) folgt aus (b) und den Vorüberlegungen 2.4.2.1. - 2.4.2.3.

Ein handliches, schon bei Huber (1964, S. 97) auftretendes Kriterium für die Existenz und Eindeutigkeit liefert Teil (ii) des folgenden Satzes:

2.4.2.9. Satz: Es sei P ein W -Maß auf $(\mathbb{R}, \mathcal{L})$. Dann gilt:

(i) Ist $P(\{a\}) < 1 - \frac{1}{k^2} \forall a \in \mathbb{R}$, so existiert zu jedem $m \in \mathbb{R}$

genau ein $v(m)$ mit $\int (x-m)^2 dP_{m, k\sqrt{v(m)}}(x) = v(m)$. Die Abbildung $m \mapsto v(m)$ ist stetig. Ferner ist die Funktion f mit

$$f(m) := \int x dP_{m, k\sqrt{v(m)}}(x) - m \quad (22)$$

stetig, und es gilt: $\lim_{m \rightarrow -\infty} f(m) = \infty$ und $\lim_{m \rightarrow \infty} f(m) = -\infty$.

(ii) Ist $P(\{a\}) < 1 - \frac{1}{k^2} \forall a \in \mathbb{R}$, so existiert genau ein regulärer Fixpunkt von P .

(iii) Existiert ein regulärer Fixpunkt von P , so besitzt P keinen weiteren (regulären oder singulären) Fixpunkt.

Beweis: (i) Die Wohldefiniertheit bzw. Stetigkeit von $m \mapsto v(m)$ folgt aus 2.3. bzw. ist offensichtlich. Hieraus folgt die Stetigkeit von f . Ferner ist klar, daß $\lim_{m \rightarrow -\infty} v(m) = \lim_{m \rightarrow \infty} v(m) = \infty$. Für $m \rightarrow \infty$ befindet sich schließlich ein beliebig großer Anteil der Wahrscheinlichkeitsmasse von P in einer im Verhältnis zu $v(m)$ beliebig kleinen Umgebung von $m - \sqrt{v(m)}$. Also folgt:

$$\lim_{m \rightarrow \infty} f(m) = \lim_{m \rightarrow \infty} [1 \cdot (m - \sqrt{v(m)}) - m] = -\lim_{m \rightarrow \infty} \sqrt{v(m)} = -\infty.$$

Entsprechend ergibt sich $\lim_{m \rightarrow -\infty} f(m) = \infty$.

(ii) Es sei Q die Standardnormalverteilung und P zunächst ein Maß mit $E_P(X^2) < \infty$ und $P(\{a\}) < 1 - \frac{1}{k^2} \forall a \in \mathbb{R}$. Dann gilt für jeden Fixpunkt (m, v) eines jeden Maßes $t \cdot P + (1-t) \cdot Q$ mit $t \in [0, 1]$ einerseits

$$\begin{aligned} v &= \int (x-m)^2 d(t \cdot P + (1-t) \cdot Q)_{m, k\sqrt{v}}(x) \leq \text{Max}(E_P(X^2), E_Q(X^2)) \\ &=: c_2 < \infty \end{aligned}$$

und andererseits

$$v > c_1 := \inf \left\{ c > 0 \mid \exists t \in [0,1] \exists a \in \mathbb{R} : t \cdot P[a-c, a+c] + (1-t) \cdot Q[a-c, a+c] > 1 - \frac{1}{k^2} \right\} > 0.$$

(Dabei ist $c_1 > 0$, da sonst für ein $a_0 \in \mathbb{R}$ $P(\{a_0\}) \geq 1 - \frac{1}{k^2}$ folgen würde.) Bezeichnet $P^{-1}(\alpha)$ das α -Fraktile von P , so gilt ferner

$$m - k\sqrt{v} < \max \left\{ P^{-1} \left[1 - \frac{1}{k^2} \right], Q^{-1} \left[1 - \frac{1}{k^2} \right] \right\} =: a_1$$

$$m + k\sqrt{v} > \min \left\{ P^{-1} \left[\frac{1}{k^2} \right], Q^{-1} \left[\frac{1}{k^2} \right] \right\} =: a_2.$$

Die kompakte Menge $[b_1, b_2] \times [c_1, c_2]$ mit $b_1 := a_2 - k\sqrt{c_2}$ und $b_2 := a_1 + k\sqrt{c_2}$ leistet daher das in 2.4.2.8.(a) Verlangte. 2.4.2.8.(b) ergibt sich aus 2.4.2.1. - 2.4.2.3., so daß die Existenz und Eindeutigkeit eines regulären Fixpunktes aus 2.4.2.8. folgt. Ist schließlich $E_p(X^2) = \infty$, so ergibt sich die Existenz direkt aus (i) mit dem Zwischenwertsatz. Der Beweis der Eindeutigkeit kann auf den Fall $E_p(X^2) < \infty$ zurückgeführt werden.

(iii) P besitze die beiden Fixpunkte (m_1, v_1) und (m_2, v_2) . O.B.d.A.

kann man annehmen, daß $(m_1, v_1) = \left[0, \frac{1}{k^2} \right]$, $v_2 > \frac{1}{k^2}$ und somit nach

dem Schachtelungssatz 2.4.2.6. $[-1, 1] \subset [m_2 - k\sqrt{v_2}, m_2 + k\sqrt{v_2}]$ ist und daß $P([m_2 - k\sqrt{v_2}, m_2 + k\sqrt{v_2}]) = 1$. Weil (m_2, v_2) ein regulärer Fixpunkt sein soll, kann man weiter o.B.d.A. davon ausgehen, daß $P(\{m_2 - k\sqrt{v_2}\}) \neq P([m_2 - k\sqrt{v_2}, -1]) > 0$. (Andernfalls kann man direkt so fortfahren wie weiter unten hinter "Fall II" ausgeführt wird.) Wir verändern P über dem Intervall $[m_2 - k\sqrt{v_2}, -1]$ in folgender Weise:

Fall I: Es existiere ein $a \in]m_2 - k\sqrt{v_2}, -1[$ mit $P(\{a\}) = P([m_2 - k\sqrt{v_2}, -1])$. Dann ersetze man P über dem Intervall $[m_2 - k\sqrt{v_2}, -1]$ durch das Maß mit der Lebesgue-Dichte $\frac{1}{2\varepsilon} \cdot P(\{a\}) \cdot 1_{[a-\varepsilon, a+\varepsilon]}$, wobei $\varepsilon := \frac{1}{4} \cdot \min\{1, |a - m_2 + k\sqrt{v_2}|, |a+1|\}$. Das entstehende Maß hat dann die Fixpunkte $\left[0, \frac{1}{k^2} \right]$ und (m_2, v'_2) mit $v'_2 = v_2 + \frac{1}{3} \cdot P(\{a\}) \cdot \varepsilon^2$. Man verfähre für diese neue Situation weiter wie im Fall II.

Fall II: Es existiert kein $a \in]m_2 - k\sqrt{v_2}, -1[$ mit $P(\{a\}) = P([m_2 - k\sqrt{v_2}, -1[)$. Dann liegt der Vektor der ersten drei zentralen Momente von P über $[m_2 - k\sqrt{v_2}, -1[$ im Inneren des entsprechenden Momentenraumes M_3 über $[m_2 - k\sqrt{v_2}, -1[$. Ohne Veränderung dieser Momente (und daher der Fixpunkte) kann man somit P über $[m_2 - k\sqrt{v_2}, -1[$ durch das zu der oberen Hauptdarstellung ("upper principal representation", vgl. z.B. Karlin/Studden 1966, S. 46 oder Krein/Nudel'man 1977, S. 78) gehörende Maß H ersetzen, das alle Masse auf zwei Punkte $m_2 \cdot k\sqrt{v_2} < s_1 < s_2 = -1$ konzentriert. Dabei ist $H(\{s_1\}) > 0$ und $H(\{s_2\}) > 0$. Aus Stetigkeitsgründen liegt daher für hinreichend kleines $\varepsilon > 0$ auch der Vektor der ersten drei zentralen Momente des Maßes $H(\{s_1\}) \cdot \delta_{s_1} + (H(\{s_2\}) - \varepsilon) \cdot \delta_{s_2}$ im Inneren von M_3 . Nach 2.4.2.4. (oder 2.4.2.5.) kann daher dieses Maß ohne Abänderung dieser Momente durch ein in $[m_2 - k\sqrt{v_2}, -1]$ atomloses Maß ersetzt werden.

Nach Konstruktion liegt im Intervall $[-1, 1]$ nunmehr positive Masse mindestens auf -1 und einem Punkt $a \in]-1, 1[$. Wiederum nach 2.4.2.4. (oder 2.4.2.5.) kann man daher zu einem über $[-1, 1]$ atomlosen Maß übergehen. Schließlich ändere man P über $]1, m_2 + k\sqrt{v_2}]$ wie folgt ab:

1. Fall: Es ist $P([1, m_2 + k\sqrt{v_2}]) = 0$. Dann bleibt P dort unverändert.
2. Fall: Es ist $P(\{m_2 + k\sqrt{v_2}\}) = P([1, m_2 + k\sqrt{v_2}]) > 0$. Dann lege man die Masse $P(\{m_2 + k\sqrt{v_2}\})$ gleichverteilt im Intervall $[m_2 + k\sqrt{v_2}, m_2 + k\sqrt{v_2} + 1]$ ab.
3. Fall: Es existiert ein $a \in]1, m_2 + k\sqrt{v_2}[$ mit $P(\{a\}) = P([1, m_2 + k\sqrt{v_2}]) > 0$. Dann verfähre man wie im Fall I für das Intervall $[m_2 - k\sqrt{v_2}, -1[$. Die neuen Fixpunkte sind $\left[0, \frac{1}{k^2}\right]$ und (m_2, v_2'') mit $v_2'' > v_2'$.
4. Fall: In den verbleibenden Fällen kann man wieder nach 2.4.2.4. (oder 2.4.2.5.) zu einem atomlosen Maß mit den gleichen ersten drei zentralen Momenten übergehen.

Das so konstruierte Maß ist atomlos und besitzt zwei voneinander verschiedene Fixpunkte. Dies widerspricht (ii).

2.4.2.10. Bemerkung:

- (i) Der hier gewählte Beweisgang über eine Homotopieüberlegung mag für diesen Spezialfall unnötig kompliziert erscheinen. Andererseits liefert er aber i.w. auf einen Schlag die Existenz und Eindeutigkeit, eine exakte Behandlung von diskreten wie stetigen Verteilungen und (zumindest für den Fall atomloser Maße) einen Algorithmus zur Berechnung (vgl. 1.3.3.).
- (ii) Über die Menge der Fixpunkte eines beliebigen Maßes P auf $(\mathbb{R}, \mathcal{L})$ läßt sich über (iii) hinaus wenig sagen. Das Beispiel $P(\{-1\}) = P(\{1\}) = \frac{1}{4k^2}$ und $P(\{0\}) = 1 - \frac{1}{2k^2}$ zeigt, daß überhaupt kein Fixpunkt zu existieren braucht, während das Beispiel $Q(\{-1\}) = Q(\{1\}) = \frac{1}{2k^2}$ und $Q(\{0\}) = 1 - \frac{1}{k^2}$ unendlich viele (singuläre) Fixpunkte besitzt. Ganz allgemein gilt aber, daß die Menge der Fixpunkte von P stets beschränkt ist. Dies folgt sofort aus 2.4.2.9.(iii).
- (iii) Die entscheidende Bedingung für die Eindeutigkeit ist die Regularität des Fixpunktes und nicht so sehr die "Hyper-ebenenbedingung" $P(\{a\}) < 1 - \frac{1}{k^2} \forall a \in \mathbb{R}$. Letztere ist lediglich die einfachste hinreichende Bedingung dafür, daß das Maß P im Intervall $]m - k\sqrt{v}, m + k\sqrt{v}[$ nicht zu einem Einpunktmaß degeneriert ist. Für die praktische Anwendung ist dies zwar belanglos, in struktureller Hinsicht (im Sinne der Äquivalenzklasseninterpretation) aber von Interesse.

2.4.3. Beschreibung von $d^r(\cdot)$:

Nach 2.4.2.9. ist auf der Menge

$$\mathcal{P}_0 := \{W\text{-Maße auf } (\mathbb{R}, \mathcal{L}) \mid \exists \text{ ein regulärer Fixpunkt } (m, v)\}$$

das robustifizierte Funktional $d^r(\cdot) = (m(\cdot), v(\cdot))^T$ erklärt. Eine besonders wichtige Teilmenge von $\mathcal{P}_0$ ist

$$\mathcal{P}'_0 := \left\{ W\text{-Maße auf } (\mathbb{R}, \mathcal{L}) \mid P(\{a\}) < 1 - \frac{1}{k^2} \forall a \in \mathbb{R} \right\}.$$

Wir stellen einige Eigenschaften von $m(\cdot)$ und $v(\cdot)$ zusammen.

2.4.3.1. Satz: Die Funktionale

$$m : \mathcal{P}_0 \longrightarrow \mathbb{R} \text{ mit } P \longmapsto m(P)$$

und

$$v : \mathcal{P}_0 \longrightarrow]0, \infty[\text{ mit } P \longmapsto v(P)$$

sind stetig. Dabei sei $\mathbb{R}$ bzw. $]0, \infty[$ mit der euklidischen und $\mathcal{P}_0$ mit der Relativtopologie der schwachen Topologie versehen.

Beweis: Es sei $P_0 \in \mathcal{P}_0$ ein fest gewähltes Maß, $m_0 := m(P_0)$, $v_0 := v(P_0)$ und o.B.d.A. $(m_0, v_0) = \left[0, \frac{1}{k^2}\right]$. Ferner bezeichne zur Abkürzung für jedes $m \in \mathbb{R}$ und $v \in]0, \infty[$ $f_{m,v}$ die durch

$$f_{m,v}(x) = \begin{cases} (x-m)^2 & , \quad |x-m| < kv\sqrt{v} \\ k^2 v & , \quad |x-m| \geq kv\sqrt{v} \end{cases} \quad (23)$$

und $g_{m,v}$ die durch

$$g_{m,v}(x) = \begin{cases} m - kv\sqrt{v} & , \quad x < m - kv\sqrt{v} \\ x & , \quad x \in [m - kv\sqrt{v}, m + kv\sqrt{v}] \\ m + kv\sqrt{v} & , \quad x > m + kv\sqrt{v} \end{cases} \quad (24)$$

gegebene Funktion.

(1) Da (m_0, v_0) ein regulärer Fixpunkt ist und P_0 nur abzählbar viele Atome besitzt, gibt es ein $a > 0$ derart, daß für $A := [m_0 - a, m_0 + a]$ gilt: $P_0(A) < 1 - \frac{1}{k^2}$ und $P_0(\partial A) = 0$. Es gibt daher auch eine Umgebung $\mathcal{U}_1(P_0) \subset \mathcal{P}_0$ derart, daß für alle $Q \in \mathcal{U}_1(P_0)$ die Beziehung $Q(A) < 1 - \frac{1}{k^2}$ gilt. Auf A ist daher für jedes $Q \in \mathcal{U}_1(P_0)$ eine Abbildung $\tilde{v}_Q : m \mapsto \tilde{v}_Q(m)$ durch die Vorschrift

$$\int f_{m, \tilde{v}_Q(m)}(x) dQ(x) = \tilde{v}_Q(m)$$

erklärt (denn die Existenz ist gesichert, und die Wohldefiniertheit folgt aus 2.3.). $\tilde{v}_Q(\cdot)$ ist für jedes $Q \in \mathcal{U}_1(P_0)$ Lipschitz-stetig. Ferner ist für jedes $m \in A$ die Abbildung $F_m : \mathcal{P}_0 \mapsto \mathbb{R}$ mit

$$Q \mapsto \int g_{m, \tilde{v}_Q(m)}(x) dQ(x) - m$$

erklärt und stetig. Für das feste P_0 ist $F(P_0)$ als Funktion von m Lipschitz-stetig und als Komposition in der Form

$$m \mapsto \left[\begin{matrix} m \\ \tilde{v}_{P_0}(m) \end{matrix} \right] \mapsto d((m, \tilde{v}_{P_0}(m)), P_0) \mapsto F_{P_0}(P_0)$$

mit d aus (11) darstellbar. Damit erhält man unter Benutzung der Ergebnisse aus 2.4.2.3.(a)

$$\partial F_{m_0}(P_0) = \text{co} \left\{ \frac{1}{d_{11}^{22}} \cdot \det(D_1), i = 1, \dots, 4 \right\},$$

wobei D_1 die Matrizen aus 2.4.2.3.(a) und d_{11}^{22} jeweils das rechte untere Element bezeichnen. Weil (m_0, v_0) ein regulärer Fixpunkt ist, ist jedes Element von $\partial F_{m_0}(P_0)$ negativ; mithin folgt aus F.H.

Clarke (1983, Theorem 2.3.7., S. 41), daß

$$F_m(P_0) > 0 \quad \forall m \in [m_0 - a, m_0[\quad \text{und} \quad F_m(P_0) < 0 \quad \forall m \in]m_0, m_0 + a].$$

(2) Für ein festes $m \in A$ gilt mit den Abkürzungen $c_p := \tilde{v}_p(m)$ und

$$c_0 := \tilde{v}_0(m) \quad \text{zunächst} \quad P_0([m - k\sqrt{c_p}, m + k\sqrt{c_p}]) = 1 - \frac{1}{k^2} + r \quad \text{mit} \quad r > 0.$$

Für $c'_1 := \inf \left\{ c \mid P_0([m - k\sqrt{c}, m + k\sqrt{c}]) \geq 1 - \frac{1}{k^2} + \frac{r}{2} \right\}$ gilt somit $c'_1 < c_p$,

denn wäre $c'_1 = c_p$, so müßte für jedes $c < c_p$ $P_0([m - k\sqrt{c}, m + k\sqrt{c}])$

$$< 1 - \frac{1}{k^2} + \frac{r}{2} \quad \text{und daher auch} \quad P_0([m - k\sqrt{c_p}, m + k\sqrt{c_p}]) \leq 1 - \frac{1}{k^2} + \frac{r}{2} \quad \text{sein.}$$

Man kann also ein $c_1 \in]c'_1, c_p[$ wählen mit $P_0([m - k\sqrt{c_1}, m + k\sqrt{c_1}])$

$$> 1 - \frac{1}{k^2} + \frac{r}{2} \quad \text{und} \quad P_0(\{m - k\sqrt{c_1}, m + k\sqrt{c_1}\}) = 0. \quad \text{Dann gilt} \quad \int f_{m, c_1}(x) dP_0(x)$$

$$= c_1 + b \quad \text{mit} \quad b > 0, \quad \text{und man kann eine Umgebung } \mathcal{U}_2(P_0) \text{ finden mit}$$

$$\int f_{m, c}(x) dQ(x) > c_1 + \frac{b}{2} \quad \forall Q \in \mathcal{U}_2(P_0). \quad \text{Für jedes } Q \in \mathcal{U}_2(P_0) \text{ ist daher}$$

$$c_Q > c_1. \quad \text{Ferner kann man eine Umgebung } \mathcal{U}_3(P_0) \text{ angeben derart, daß}$$

$$|P_0([m - k\sqrt{c_1}, m + k\sqrt{c_1}]) - Q([m - k\sqrt{c_1}, m + k\sqrt{c_1}])| < \frac{r}{4} \quad \forall Q \in \mathcal{U}_3(P_0). \quad \text{Für}$$

jedes $Q \in \mathcal{U}_2(P_0) \cap \mathcal{U}_3(P_0)$ ist dann

$$Q([m - k\sqrt{c_Q}, m + k\sqrt{c_Q}]) > Q([m - k\sqrt{c_1}, m + k\sqrt{c_1}]) \geq 1 - \frac{1}{k^2} + \frac{r}{4}.$$

Schließlich gibt es eine Umgebung $\mathcal{U}_4(P_0)$, so daß $\forall Q \in \mathcal{U}_4(P_0)$

$$\left| \int f_{m, c_p}(x) dP_0(x) - \int f_{m, c_p}(x) dQ(x) \right| < \varepsilon.$$

Für ein beliebig vorgegebenes $\varepsilon > 0$ und $Q \in \bigcap_{i=1}^4 \mathcal{U}_i(P)$ gilt:

$$|c_p - c_Q| = \left| \int f_{m, c_p} dP_0 - \int f_{m, c_Q} dQ \right|$$

$$\leq \left| \int f_{m, c_p} dP_0 - \int f_{m, c_p} dQ \right| + \left| \int f_{m, c_p} dQ - \int f_{m, c_Q} dQ \right|$$

$$\leq \varepsilon + |c_p - c_q| \cdot k^2 \cdot (1 - Q([m - k\sqrt{c_1}, m + k\sqrt{c_1}])).$$

Also:

$$|c_p - c_q| \leq \varepsilon \cdot \frac{4}{rk^2}.$$

(3) Zu vorgegebenem $m \in A$ und $\varepsilon > 0$ lässt sich nach (1) und (2) eine Umgebung $\mathcal{U}_5(P_0)$ mit $|c_p - c_q| < \frac{\varepsilon}{2} \cdot \frac{1}{k} \quad \forall Q \in \mathcal{U}_5(P_0)$ und eine Umgebung $\mathcal{U}_6(P_0)$ mit $|\int g_{m,c_p} dP - \int g_{m,c_p} dQ| < \frac{\varepsilon}{2} \quad \forall Q \in \mathcal{U}_6(P_0)$ finden. Für $Q \in \mathcal{U}_5(P_0) \cap \mathcal{U}_6(P_0)$ gilt dann:

$$\begin{aligned} |F_m(P_0) - F_m(Q)| &= |\int g_{m,c_p} dP_0 - \int g_{m,c_q} dQ| \\ &\leq |\int g_{m,c_p} dP_0 - \int g_{m,c_p} dQ| + |\int g_{m,c_p} dQ - \int g_{m,c_q} dQ| \\ &\leq \frac{\varepsilon}{2} + |c_p - c_q| \cdot k \\ &\leq \varepsilon. \end{aligned}$$

Damit ist gezeigt: Für jedes $m \in A$ ist $F_m(\cdot)$ bei P_0 stetig.

(4) Für hinreichend kleines $\varepsilon > 0$ ist $F_{m_0 - \varepsilon}(P_0) > 0$ und $F_{m_0 + \varepsilon}(P_0) < 0$. Also existieren Umgebungen $\mathcal{U}_7(P_0)$ mit $F_{m_0 - \varepsilon}(Q) > 0 \quad \forall Q \in \mathcal{U}_7(P_0)$ und $\mathcal{U}_8(P_0)$ mit $F_{m_0 + \varepsilon}(Q) < 0 \quad \forall Q \in \mathcal{U}_8(P_0)$. Für alle $Q \in \mathcal{U}_7(P_0) \cap \mathcal{U}_8(P_0)$ folgt also $m(Q) \in [m_0 - \varepsilon, m_0 + \varepsilon]$, was die Stetigkeit von $m(\cdot)$ impliziert.

(5) Wegen

$$|\tilde{v}_{P_0}(m(P_0)) - \tilde{v}_Q(m(Q))| \leq |\tilde{v}_{P_0}(m(P_0)) - \tilde{v}_Q(m(P_0))| + |\tilde{v}_Q(m(P_0)) - \tilde{v}_Q(m(Q))|$$

folgt auch die Stetigkeit von $v(\cdot)$.

2.4.3.2. Folgerung (Konsistenz): Für jedes $P \in \mathcal{P}_0$ ist die Schätzerfolge $d^r(P^n)$ konsistent für $d^r(P)$.

Beweis: Glivenko-Cantelli.

2.4.3.3. Folgerung (Robustheit): Für jedes $F \in \mathcal{P}_0$ ist die Schätzerfolge $d^r(P^n)$ robust und Π -robust im Sinne von Hampel.

Beweis: Hampel (1971, Theorem 2 und Theorem 4, S. 1892ff).

Asymptotische Normalität kann wiederum nicht für alle $P \in \mathcal{P}_0$ gezeigt werden (vgl. 2.2.); wir beschränken uns auf den praktisch relevanten Fall, daß das zugrundeliegende W-Maß atomlos ist. P^n bezeichnet wieder die empirische Verteilung von n i.i.d. Realisationen.

2.4.3.4. Satz (asymptotische Normalität): Es sei $P \in \mathcal{P}_0$ ein atomloses W-Maß. Dann gilt:

$$\sqrt{n} \left[\begin{pmatrix} m(P^n) \\ v(P^n) \end{pmatrix} - \begin{pmatrix} m(P) \\ v(P) \end{pmatrix} \right] \xrightarrow{\mathcal{D}} N_2 \left[0, \Lambda^{-1} C (\Lambda^T)^{-1} \right]$$

mit

$$\Lambda = \begin{pmatrix} -p_2 & \frac{k^2}{2k\sqrt{v(P)}} \cdot (p_3 - p_1) \\ 2k\sqrt{v(P)} \cdot (p_3 - p_1) & k^2 - 1 - k^2 p_2 \end{pmatrix} \quad (25)$$

und

$$C = (c_{ij})_{\substack{i=1,2 \\ j=1,2}},$$

wobei

$$c_{11} = v(P)$$

$$c_{12} = c_{21} = \left[k\sqrt{v(P)} \right]^3 (p_3 - p_1) + \int_I (x-m)^3 dP(x)$$

$$c_{22} = (p_1 + p_3) \left[(k^2 - 1)^2 + 1 \right] v(P)^2 - 2v(P)^3 (1 - k^2(p_1 + p_3)) + \int_I (x-m)^4 dP(x).$$

Dabei ist

$$I :=]m(P) - k\sqrt{v(P)}, m(P) + k\sqrt{v(P)}[$$

$$p_1 := P(]-\infty, m(P) - k\sqrt{v(P)}])$$

$$p_2 := P(]m(P) - k\sqrt{v(P)}, m(P) + k\sqrt{v(P)}[)$$

$$p_3 := P([m(P) + k\sqrt{v(P)}, \infty[).$$

Ist speziell P symmetrisch (notwendigerweise um $m(P)$), so gilt:

$$\Lambda^{-1} C (\Lambda^T)^{-1} = \begin{pmatrix} \frac{v(P)}{p_2^2} & 0 \\ 0 & \frac{c_{22}}{(k^2 - 1 - k^2 p_2)^2} \end{pmatrix}. \quad (26)$$

Beweis: Die Konsistenz ist nach 2.4.3.2. sichergestellt. Man verifiziert die Bedingungen (N-1) bis (N-4) aus Huber (1981, S. 132f). Corollary 3.2 aus Huber (1981, S. 133) liefert dann das obige Resultat. Die Matrix Λ aus (25) ist bis auf die Normierung $k\sqrt{v(P)} = 1$ die aus (17) bekannte Matrix, die auch in diesem Zusammenhang wichtig ist. Die Matrix C ergibt sich als Kovarianzmatrix von

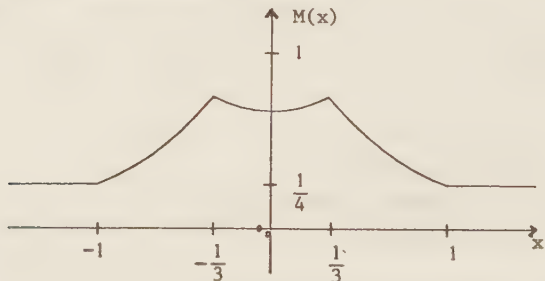
$$\psi(x, \theta_0) := \begin{pmatrix} g_{m(P), v(P)}(x) - m(P) \\ f_{m(P), v(P)}(x) - v(P) \end{pmatrix}$$

mit f und g wie in (23) und (24).

Schließlich ist $d^r(\cdot)$ lokations- und skalenequivariant (d.h. equivariant unter affinen Transformationen $x \mapsto ax + b$ mit $a > 0$) und legt in $\mathcal{P}_0$ eine Äquivalenzrelation gemäß 1.3.7. fest. Im folgenden soll für die spezielle Wahl $k := \sqrt{3}$ die Klasse $Kl(0, \frac{1}{3}) = Kl((0, \frac{1}{3}); \mathcal{P}_0)$ noch genauer beschrieben werden. Die Maximalmassenfunktion ist

$$M(x) = \begin{cases} \frac{2}{3-3x^2} & , \quad |x| < \frac{1}{3} \\ \frac{1}{3x^2+1} & , \quad \frac{1}{3} \leq |x| \leq 1 \\ \frac{1}{4} & , \quad 1 < |x| \end{cases}$$

und läßt sich folgendermaßen skizzieren:

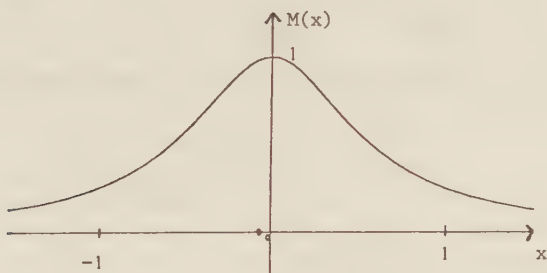


Es ist $M_o = \frac{3}{4}$, $M_u = \frac{1}{4}$, $M_B = \frac{1}{2}$. Der Breakdown-Punkt ist offenbar $M_u = \frac{1}{4}$, weil M etwa den beschränkten Träger $[-1, 1]$ besitzt. Erwartungsge-
mäß verschlechtert sich also der Breakdown-Punkt von der Lokati-
ons- über die Skalen- zur simultanen Schätzung, und zwar (als
ästhetische oder mnemotechnische Randnotiz) in der Reihenfolge $\frac{1}{2}$,

$\frac{1}{3}, \frac{1}{4}$. Bildet man die entsprechende Klasse $Kl((0, \frac{1}{3}); \mathcal{P}_2)$ in der Menge $\mathcal{P}_2$ aller W-Maße auf $(\mathbb{R}, \mathcal{L})$ mit endlichem zweiten Moment bezüglich der "unrobustifizierten" Funktionale Mittelwert und Varianz, so ergibt sich die Maximalmassenfunktion

$$M(x) = \frac{1}{3x^2 + 1},$$

also als Skizze:



Wie schon im Fall der reinen Lokations- bzw. Skalenschätzung stimmt die "robustifizierte" Maximalmassenfunktion mit der "unrobustifizierten" auf einem zentralen Bereich $a < |x| < b$ überein und wird für $|x| > b$ und gegebenenfalls $|x| < a$ abgeändert. Es ist hier $M_0=1$, $M_u=0$, $M_B=1$. Auch diese Funktion hat etwa den beschränkten Träger $[-1, 1]$; somit ist der Breakdown-Punkt 0. Als Momentenproblem betrachtet, ergibt sich

- für Mittelwert und Varianz:

$$Kl(0, \frac{1}{3}) = \left\{ P \in \mathcal{P}_2 \mid P \text{ löst das Hamburger-Momentenproblem} \right. \\ \left. \text{mit } \mu_0=1, \mu_1=0, \mu_2=\frac{1}{3} \right\}$$

- für $d^r(\cdot)$:

$$Kl(0, \frac{1}{3}) = \left\{ P \in \mathcal{P}_0 \mid \text{Die Einschränkung von } P \text{ auf } [-1, 1] \text{ ist} \right. \\ \text{eine Lösung des Stieltjes-Momentenproblems auf} \\ \left. [-1, 1] \text{ mit } \mu_0=1-p_1-p_3, \mu_1=p_1-p_3 \text{ und } \mu_2=\frac{1}{3}-p_1-p_3, \right. \\ \left. \text{wobei } p_1=P((-\infty, -1]) \text{ und } p_3=P([1, \infty)) \right\}.$$

Die Zahlen p_1 und p_3 beschreiben wieder die Klasse $Kl(0, \frac{1}{3})$ genauer. In diesem Fall ist es jedoch zweckmäßiger, mit den Bezeichnungen aus (16) die Zahlen $p'_1 := p_1 + r_1$ und $p'_3 := p_3 + r_3$ als "Kennzahlen" zu wählen, weil dann auch die Bedingung " $P \in \mathcal{P}_0$ " durch die Kennzahlen ausgedrückt werden kann. Es ist bequem, auch

noch die Koordinaten $x := p'_1 - p'_3$ und $y := p'_1 + p'_3$ einzuführen.

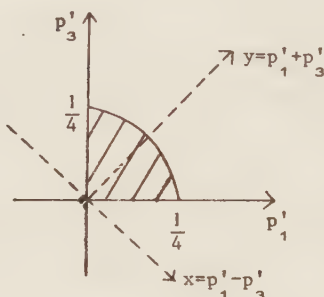
Unter Benutzung der Determinante $\tilde{D}$ aus (21) ergeben sich dann die Äquivalenzen

$$P \in \mathcal{P}_0 \Leftrightarrow (1-p'_1-p'_3)\left(\frac{1}{3}-p'_1-p'_3\right) - (p'_1-p'_3)^2 > 0$$

$$\Leftrightarrow (1-y)\left(\frac{1}{3}-y\right) - x^2 > 0$$

$$\Leftrightarrow p'_1 < \frac{1}{4} \cdot \frac{1-4p'_3}{1-3p'_3}.$$

Der entstehende Kennzahlenbereich ist in der folgenden Skizze dargestellt:



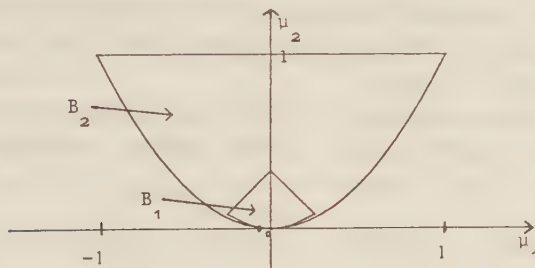
Wir notieren drei Anmerkungen:

- (i) Der Kennzahlbereich ist konvex. Dies folgt aus der Konvexität von $\mathcal{P}_0$ und der Tatsache, daß p'_1 und p'_3 Linearformen auf

$$\mathcal{P}_0 \text{ sind oder -elementar- wegen } \left[\frac{1}{4} \cdot \frac{1-4p'_3}{1-3p'_3} \right]'' = -\frac{3}{2} \cdot \frac{1}{(1-3p'_3)^3} < 0.$$

- (ii) Das Minimum und Maximum der Determinante (21) wird in Extrempunkten des Kennzahlbereichs angenommen. Weil die Determinante (21) aber keine konkave Funktion in p'_1 und p'_3 ist (z.B. ergibt sich für $p'_1 = p'_3 =: p$ die über $[0, \frac{1}{6}]$ konvexe Funktion $\frac{1}{3} - \frac{8}{3}p + 4p^2$), kann dies nicht - etwa via H. Bauers "Minimumprinzip" (Bauer 1958, S. 392) - a priori zur Beschränkung auf extremale Maße genutzt werden.
- (iii) Die Determinante (21) nimmt über den Kennzahlenbereich ihr ganzes Wertespektrum bereits für "symmetrische" Maße (d.h. für den Fall $p'_1 = p'_3$) an. Auch diese Beobachtung läßt sich aber nicht a priori beweisen (und zur Vereinfachung des Beweises der Regularität nutzen).

Die ersten und zweiten Momente über $[-1,1]$ eines Maßes P aus $\mathcal{P}_0$ sind $\mu_1 = x = p'_1 - p'_3$ und $\mu_2 = \frac{1}{3} - y = \frac{1}{3} - p'_1 - p'_3$. Der durch die Maße aus $Kl((0, \frac{1}{3}); \mathcal{P}_0)$ abgedeckte Bereich B_1 von (μ_1, μ_2) -Werten und der entsprechende Bereich B_2 für beliebige substochastische Maße sind in der folgenden Skizze wiedergegeben:



Setzt man $f_1(x) := g_{0, \frac{1}{3}}(x)$ und $f_2(x) := f_{0, \frac{1}{3}}(x) - \frac{1}{3}$ mit den Funktionen aus (23) und (24), so ergeben sich als Extrempunkte für die Menge der W -Maße P mit $\int f_1(x) dP(x) = 0$ und $\int f_2(x) dP(x) = 0$ nach 1.3.8.2. die folgenden Maße:

- die Zwei-Punkt-Maße $P = p_1 \delta_{x_1} + p_2 \delta_{x_2}$ mit

$$(i) \quad p_2 \in]\frac{1}{4}, \frac{3}{4}[\text{ und } p_1 = 1 - p_2, \quad x_2 = \sqrt{\frac{1-p_2}{p_2}} \cdot \sqrt{\frac{1}{3}}, \quad x_1 = -\frac{1}{3x_2}$$

oder

$$(ii) \quad p_2 = \frac{1}{4} \text{ und } p_1 = \frac{3}{4}, \quad x_2 \in [1, \infty[, \quad x_1 = \frac{1}{3}$$

oder

$$(iii) \quad p_2 = \frac{3}{4} \text{ und } p_1 = \frac{1}{4}, \quad x_2 = \frac{1}{3}, \quad x_1 \in]-\infty, -1]$$

- die Drei-Punkt-Maße $P = p_1 \delta_{x_2} + p_2 \delta_{x_2} + p_3 \delta_{x_3}$ mit $p_1 > 0$, $\sum p_i = 1$, $\sum p_i x_i = 0$, $\sum p_i x_i^2 = \frac{1}{3}$ und

$$\det \begin{pmatrix} 1 & f_1(x_1) & f_2(x_1) \\ 1 & f_1(x_2) & f_2(x_2) \\ 1 & f_1(x_3) & f_2(x_3) \end{pmatrix} \neq 0.$$

Zur Sicherung der Eindeutigkeit müssen zur Bildung der Klasse $Kl((0, \frac{1}{3}); \mathcal{P}_0)$ einige "Randstücke" (die auch zu anderen Klassen gehören würden) ausgeschlossen werden. Aus diesem Grunde ist es auch nicht einfach, die Separationstheorie für konvexe Mengen

direkt zum Beweis der Eindeutigkeit der Schätzungen heranzuziehen.
In diesem Zusammenhang sind folgende Anmerkungen von Interesse:

- (i) Hat das Maß P den Fixpunkt $(0, \frac{1}{3})$ und gibt es ein $x_0 \in \mathbb{R}$ mit $P(\{x_0\}) = M(x_0)$ (d.h. P legt die Maximalmasse auf die Stelle x_0), so ist $(0, \frac{1}{3})$ singulärer Fixpunkt von P . P besitzt dann keinen regulären Fixpunkt und liegt nicht in $Kl((0, \frac{1}{3}); \mathcal{P}_0)$. Die Maße mit Maximalmasse liegen insofern am "Rande" der Klasse $Kl((0, \frac{1}{3}); \mathcal{P}_0)$.
- (ii) Ein Maß P , das auf eine Stelle x_0 die Maximalmasse $M(x_0)$ legt, braucht nicht notwendigerweise Extrempunkt von $Kl((0, \frac{1}{3}); \mathcal{P}_0)$ zu sein.

Beispiel: P mit $P(\{\frac{1}{3}\}) = \frac{3}{4}$ und der Lebesgue-Dichte $\frac{1}{4}$ auf dem Intervall $[1, 2]$.

- (iii) Ein Extrempunkt von $Kl((0, \frac{1}{3}); \mathcal{P}_0)$ legt nicht notwendigerweise Maximalmasse auf eine Stelle $x_0 \in \mathbb{R}$.

Beispiel: P mit $P(\{-\frac{1}{2}\}) = \frac{2}{9} < \frac{4}{7} = M(-\frac{1}{2})$, $P(\{-\frac{1}{4}\}) = \frac{8}{15} < \frac{32}{45} = M(-\frac{1}{4})$ und $P(\{1\}) = \frac{11}{45} < \frac{1}{4} = M(1)$.

2.4.4. Rechentechnische Anmerkungen:

2.4.4.1. Situation: Wir betrachten nur den Fall $k = \sqrt{3}$ und setzen zur Abkürzung $c := \sqrt{3} \sqrt{v}$. Ferner sei in diesem Abschnitt P stets ein W -Maß auf $(\mathbb{R}, \mathcal{B})$ mit $P(\{a\}) < \frac{2}{3} \forall a \in \mathbb{R}$. Dann handelt es sich darum, den (eindeutig bestimmten) Fixpunkt von

$$\int x dP_{m,c}(x) = m \quad (27)$$

$$\int (x - m)^2 dP_{m,c}(x) = \frac{c^2}{3} \quad (28)$$

zu berechnen. Nach 2.4.2.9.(i) ist die Abbildung

$$c: m \mapsto c(m) \text{ derart, daß (28) gilt} \quad (29)$$

wohldefiniert und stetig; man kann also statt des obigen Fixpunktes auch den Fixpunkt von

$$F: m \mapsto \int x dP_{m,c(m)}(x) \quad (30)$$

oder die Nullstelle von

$$G : m \mapsto \int x dP_{m, c(m)}(x) - m \quad (31)$$

berechnen.

2.4.4.2. Bemerkung (Prinzipielle Berechenbarkeit):

- (i) Für ein diskretes W-Maß P läßt sich der Fixpunkt von (27) und (28) in endlich vielen Schritten exakt ermitteln, vgl. Huber (1964, S. 99). Dies ist ein Beispiel für den in 1.3.3.4. erwähnten Fall.
- (ii) Für jedes (stetige oder diskrete) W-Maß kann man prinzipiell die Nullstelle von (31) mit vorgegebener Genauigkeit durch ein Einschachtelungsverfahren bestimmen.

Für den Fall der eindimensionalen Funktion F aus (30) gibt der folgende Satz eine hinreichende Bedingung dafür, daß F selbst kontraktiv ist:

2.4.4.3. Satz: Die Abbildung c aus (29) genüge einer Lipschitz-Bedingung mit den Konstanten α . Dann gilt:

- (i) Falls $\alpha \leq 2$, so ist G aus (31) streng monoton fallend.
- (ii) Falls $\alpha \leq 1$, so ist F aus (30) kontraktiv.

Beweis: (i) Es sei $m_1 < m_2$, $c_1 := c(m_1)$, $c_2 := c(m_2)$ und o.B.d.A. $c_2 \geq c_1$ (sonst spiegele man P um 0).

1. Fall: Es sei $m_2 - c_2 < m_1 - c_1$. Dann folgt:

$$\begin{aligned} F(m_2) &\leq F(m_1) + P([m_1 + c_1, \infty]) \cdot (m_2 + c_2 - m_1 - c_1) \\ &< F(m_1) + \frac{1}{3} \cdot 3 \cdot (m_2 - m_1). \end{aligned}$$

2. Fall: Es sei $m_2 - c_2 \geq m_1 - c_1$ (d.h. $m_2 - m_1 \geq c_2 - c_1$). Dann folgt:

$$\begin{aligned} F(m_2) &\leq F(m_1) + P(]-\infty, m_2 - c_2]) \cdot (m_2 - c_2 - m_1 + c_1) \\ &\quad + P([m_1 + c_1, \infty]) \cdot (m_2 + c_2 - m_1 - c_1) \\ &< F(m_1) + \frac{1}{3} \cdot (m_2 - m_1) + \frac{1}{3} \cdot 2 \cdot (m_2 - m_1). \end{aligned}$$

In beiden Fällen gilt also

$$F(m_2) < F(m_1) + (m_2 - m_1). \quad (32)$$

Hieraus folgt (i).

(ii) Unter den gleichen Vereinfachungen wie in (i) gilt nun $m_2 - c_2 \geq m_1 - c_1$ und daher $F(m_2) \geq F(m_1)$. Daher folgt aus (32) sogar

$$|F(m_2) - F(m_1)| < |m_2 - m_1|,$$

und somit (ii).

2.4.4.4. Zusatz:

- (i) Für kontraktive Abbildungen und die Konvergenz der einfachen Iteration ("sukzessive Approximation") gelten ähnliche Resultate wie im (schärferen) Banachschen Fixpunktsatz, vgl. Jeggle (1979, S. 45ff).
- (ii) Die Aussagen von 2.4.4.3. gelten nicht allgemein. Als Gegenbeispiel betrachte man etwa das durch $P(\{-1000\}) = \frac{99}{300}$, $P(\{0\}) = \frac{199}{300}$ und $P(\{1\}) = \frac{2}{300}$ gegebene Maß. Dann ist $c(-1) = \sqrt{207}$, $c(0) = \sqrt{2}$, $F(-1) = -3.07$, $F(0) = -0.46$, $G(-1) = -4.07$, $G(0) = -0.46$.

2.4.4.5. Notiz (über praktische Erfahrungen):

Programmiertechnisch scheint es am bequemsten, direkte Iteration in den Gleichungen (27) und (28) zu verwenden, da man dann die gleichen Funktionale (nur für jeweils verschiedene Daten) berechnen muß. Für die im Rahmen dieser Arbeit angefallenen Berechnungen für diskrete Maße konvergierte dieses Verfahren gut (Startwerte: Mittelwert und Varianz). Für stetige Verteilungen hat man häufig Zusatzinformationen (z.B. Symmetrie). Geschicktere Algorithmen ergeben sich durch kombinatorische Überlegungen oder Hin- und Herpendeln zwischen (27) und (28) (vgl. auch das sogenannte Cobweb-Theorem in der Volkswirtschaftslehre).

2.5. Anmerkungen zur Wahl des Schätzers

In diesem Abschnitt soll auf einige Güteeigenschaften des hier gewählten Ansatzes eingegangen werden. Dazu gehen wir von den folgenden (notwendigerweise sehr subjektiven) Eindrücken aus:

- (i) Eine "zwingende" (d.h. für alle einsehbare und verbindliche) Theorie der Robustheit existiert derzeit nicht und ist aus prinzipiellen Gründen auch nicht zu erwarten (vgl. auch Bickel 1976, S.145f und Hampel e.a. 1986, S. 44f).
- (ii) Die wesentlichen Ansätze sind P.J. Hubers Minimax-Theorie (Huber 1981) und F.R. Hampels Influenzkurvenansatz (Hampel

e.a. 1986). Im Rahmen des Minimax-Ansatzes hat insbesondere die Maximierung der asymptotischen Varianz für verschiedene Schätzer über verschiedene Klassen von Verteilungen viel Augenmerk gefunden (in neuerer Zeit z.B. in den Arbeiten von Collins/Wiens 1985 und Wiens 1986). Dabei werden jedoch mehr oder weniger weitgehende Annahmen gemacht (z.B. $\psi \equiv 0$ außerhalb $[-k, k]$, F besitzt eine absolutstetige Dichte, F besitzt endliche Fisher-Information). Wohl nicht ganz ohne Berechtigung kritisiert daher Pfanzagl (1981, S. 4), die sich daraus ergebende Willkür. Auch der Influenzkurvensatz überwindet diese Willkür nicht (vgl. z.B. Hampel e.a. 1986, Abschnitt 4.2.).

- (iii) Die meisten "handgreiflichen" Resultate beziehen sich auf den Fall der eindimensionalen Lokationsschätzung.

Punkt (ii) sollte nicht als polemische Kritik an den obengenannten Ansätzen mißverstanden werden; es scheint vielmehr, daß man sich bis zu einem gewissen Grad zwangsläufig "willkürlich verhalten" - oder vielleicht besser: "an außermathematischen Überlegungen orientieren" - muß. Wir beginnen deswegen mit einigen heuristischen Anmerkungen.

2.5.1. Heuristische Notizen:

- (i) Hubers klassischer Lokationsschätzer (1) aus 1. erweist sich in vielerlei Hinsicht als optimal (vgl. Huber 1964, Huber 1968, Hampel e.a. 1986, Abschnitt 1.3.). Es scheint daher nicht unvernünftig, Winsorisieren als Grundprinzip zu nutzen.
- (ii) Die Wahl von Ellipsoiden und - sofern keine weitere Information vorliegt - speziell von Konzentrationsellipsoiden läßt sich datenanalytisch motivieren. Ein Punktschätzer ist ja stets nur soviel wert, wie sich der Statistiker letztlich darunter vorstellen kann. Eine auf Konzentrationsellipsoiden beruhende Schätzung hat hier den Vorteil, die intuitiven Beweggründe für die Einführung von Konzentrationsellipsoiden ("Veranschaulichung" von Verteilungen mit gegebenem ersten und zweiten Moment) auszunutzen. I.w. bestehen die oben abgeleiteten Schätzer darin, das Konzentrationsellipsoid "robust" zu schätzen und daraus Lokation und Scatter abzuleiten. Allerdings entzieht sich dieses Argument einer

mathematischen Formalisierung und wird (zu Recht) von Pfanzagl (1980, S. 17) als (mathematisch) ungeeignetes Kriterium außer acht gelassen.

- (iii) Da die Konsistenzkorrektur für die Scatterschätzung nachträglich erfolgt (- im Gegensatz zu Hubers Proposal 2 (Huber 1964, S. 96), wo sie in die Schätzung mit hineingenommen wird -), lassen sich die Auswirkungen für verschiedene Annahmen über die zugrundeliegende Verteilung (z.B. Normal, t, Cauchy) stets leicht studieren. Betrachtet man sogar Quotienten (z.B. im mehrdimensionalen Fall: Heritabilitätskoeffizienten), so entfällt dieses Element von "Willkür" ganz.

2.5.2. Mathematische Anmerkungen:

In Übereinstimmung mit 2.5.1. (iii) beschränken wir uns i.w. auf den eindimensionalen Lokationsfall.

2.5.2.1. Bemerkung: Die asymptotische absolute Effizienz (im Sinne von Lehmann 1983, S. 363) des Lokationsschätzers (mit $k=\sqrt{3}$) ist für symmetrische Verteilungen $\frac{p_2^2}{I_f \cdot v(P)}$, vgl. 2.1.6. und Abschnitt 2.2. Wegen (26) aus 2.4.3.4. bleibt dies auch gültig, falls die Varianz hilfsweise mitgeschätzt wird. Quantitativ ergeben sich die in Tabelle 2.1. angegebenen Resultate.

Tabelle 2.1.: Absolute (asymptotische) Effizienzen für ausgewählte Verteilungen

Verteilung	$c=\sqrt{3 \cdot v(P)}$	p_2	$\frac{c^2}{3p_2^2}$	I_f	AAE
Normal	1.5455	0.8778	1.0333	1	0.97
T(0.01,3)	1.5603	0.8765	1.0563	0.96996	0.98
T(0.05,3)	1.6224	0.8711	1.1563	0.88326	0.98
Logistic	2.5997	0.8617	3.0340	$\frac{1}{3}$	0.99
t_5	1.7355	0.8568	1.3676	$\frac{3}{4}$	0.97
t_3	1.8832	0.8438	1.6603	$\frac{2}{3}$	0.90
Cauchy	2.9445	0.7916	4.6120	$\frac{1}{2}$	0.43
DE	1.7906	0.8331	1.5399	1	0.65

Die Werte für die Fisher-Information I_f der Tukey-Kontaminationsverteilungen $T(0.01,3)$ und $T(0.05,3)$ wurden aus Gastwirth/Cohen (1970, S. 947) übernommen. Die gefundenen Resultate erscheinen im Vergleich zu denen anderer Schätzer durchaus akzeptabel, vgl. insbesondere Lehmann (1983, Table 4.4, S. 364 und Table 6.2, S. 386).

2.5.2.2. Bemerkung: Für ein beliebiges $k \in]1, \infty[$ ergibt sich als Maximalmassenfunktion der Klasse $Kl\left[\left[0, \frac{1}{k^2}\right]; \mathcal{P}_0\right]$:

$$M(x) = \begin{cases} \frac{k^2-1}{k^2(1-x^2)} & , \quad |x| < \frac{1}{k^2} \\ \frac{k^2-1}{k^2x^2+1} & , \quad \frac{1}{k^2} \leq |x| \leq 1. \\ \frac{1}{k^2+1} & , \quad 1 < |x| \end{cases}$$

Ferner ist

$$M_o(x) = \frac{k^2}{k^2+1}$$

$$M_u(x) = \begin{cases} \frac{1}{k^2+1} & , \quad k < \sqrt{\frac{1+\sqrt{5}}{2}} \\ \frac{k^2-1}{k^2} & , \quad k \geq \sqrt{\frac{1+\sqrt{5}}{2}} \end{cases}$$

$$M_B(x) = \begin{cases} \frac{k^2-1}{k^2+1} & , \quad k < \sqrt{\frac{1+\sqrt{5}}{2}} \\ \frac{1}{k^2(k^2+1)} & , \quad k \geq \sqrt{\frac{1+\sqrt{5}}{2}} \end{cases}$$

Als Grenzfälle ergeben sich für $k=\infty$ der Fall der Schätzung durch die unrobustifizierte Funktionale Mittelwert und Varianz und für $k=1$ eine Klasse, die sich auf die Verteilungen des in 1.1.7.(i) und 2.2.2.(6) beschriebenen Typs bezieht. (Diese liegen aber nicht mehr in $\mathcal{P}_0$.) Im ersten Fall ist $M(x)$ wie in 2.4.3. angegeben; im zweiten Fall wäre (rein formal) $M(x) \equiv \frac{1}{2}$ für $|x| \geq 1$ und $M(x) \equiv 0$ für $|x| < 1$. Um einen Kompromiß zwischen diesen Extremen zu finden,

kann man etwa fordern, daß $M_0 = \frac{k^2}{k^2+1} = \frac{3}{4}$ sein soll (d.h. den Mittelwert zwischen 1 und $\frac{1}{2}$ annimmt). Dann folgt sofort die oben vorgeschlagene Konzentrationsellipsoidwahl $k=\sqrt{3}$. Dieser Vorschlag nimmt in der Tat - etwa auch in Bezug auf den Wert von $\lim_{|x| \rightarrow \infty} M(x)$ - eine Art "Mittelposition" zwischen $k=1$ und $k=\infty$ ein. Der "Sprung"

im Verhalten von $M(x)$ an der Stelle $k_0 := \sqrt{\frac{1+\sqrt{5}}{2}}$ liegt darin begründet, daß für $k < k_0$ M_u an der Stelle 0 und für $k > k_0$ M_u auf dem Bereich $\{x \mid |x| \geq 1\}$ angenommen wird. Der inhaltliche Hintergrund hierfür ist wiederum, daß für $k > k_0$ als erstes die Lokationsschätzung durch "Explosion" infolge von "outliers" zusammenbricht, während für $k < k_0$ der Breakdown-Punkt durch "Implosion" der Skalenschätzung infolge von "inliers" bestimmt wird. Eine Analyse der Maximalmassenfunktion macht hier fast zwangsläufig auf einen Umstand aufmerksam, der sonst oft übersehen wird und dessen (topologische) Definition etwas verwickelt ist, vgl. Hampel e.a. (1986, S. 97f). Will man einen möglichst großen Breakdown-Punkt erreichen und fordert man daher $M_u = \text{Max!}$, so ist $k = k_0$ zu wählen. Dasselbe Resultat ergibt sich, falls die Äquivalenzklasse möglichst stark "geglättet" werden soll in dem Sinne, daß $M_B = \text{Min!}$. Es ist dann näherungsweise $M_0 = 0.618$, $M_u = 0.382$, $M_B = 0.236$. Diese "robusteste" Variante scheint zumindest im eindimensionalen Fall übertrieben pessimistisch, was zu unnötig großen Effizienzverlusten führen würde. Die obigen Überlegungen illustrieren auch, wie man mit Hilfe von $M(x)$ zu Schätzern mit bestimmten gewünschten Eigenschaften gelangen kann. Die Maximalmassenfunktion ist diesbezüglich "dual" zur Influenzfunktion. Die Influenzfunktion gibt für eine feste Verteilung den Effekt einer (infinitesimalen) zusätzlichen Masse an der Stelle x auf den Schätzer an; die Maximalmassenfunktion beschreibt für feste Werte des Schätzers die Äquivalenzklasse der dahinterstehenden Verteilungen.

Für die folgenden Bemerkungen führen wir vorweg Bezeichnungen für die Menge der unimodalen (sowie der unimodalen symmetrischen) Verteilungen (bzw. Verteilungsfunktionen) ein. Es sei:

$\mathcal{U} := \{\text{Verteilungsfunktionen } F \mid F \text{ ist unimodal im Sinne von Karlin/Studden (1966, S. 483)}\}$

$\mathcal{F} := \{\text{Verteilungsfunktionen } F \mid F(x) = 1 - F(-x) \quad \forall x \neq 0\}$

$\mathcal{U}_{\mathcal{F}} := \mathcal{U} \cap \mathcal{F}$.

2.5.2.3. Bemerkung: Unter der Annahme, daß als Lokationsschätzer ein winsorisiertes Mittel verwendet werden soll, soll k so bestimmt werden, daß

- (i) der Schätzer möglichst robust gegen "heavy tails" wird
- (ii) die Nebenbedingung $\mathcal{U}_{\mathcal{F}} \cap \{F \mid F(k, \omega) = 0\} \neq \emptyset$ erfüllt ist (d.h. für wenigstens eine für realistisch erachtete Verteilung soll "Winsorisieren" entfallen).

Dann ergibt sich $k = \sqrt{3}$. Dies ist in diesem Sinne der kleinste akzeptable Wert.

2.5.2.4. Bemerkung: Während etwa bei Hubers Proposal 2 (Huber 1964, S. 96) eine winsorisierte Normalverteilung angepaßt wird (vgl. auch 1.1.7. (iii)), wird bei dem in 2.4. betrachteten Vorschlag eine Gleichverteilung angepaßt (womit dann auch k kanonisch festliegt). Dies läßt sich in mancherlei Hinsicht auch mathematisch motivieren:

- (i) Die Gleichverteilung ist die "ungünstigste" Verteilung im Sinne der Untersuchungen von Gray/Odell (1967) und Jacobson (1969) (vgl. auch Seaman e.a. 1985).
- (ii) Häufig interessieren Lokation und Scatter nicht an sich, sondern vielmehr ein dadurch gegebener " $m \pm s$ Bereich". Dies führt zu der Frage, für welches $F \in \mathcal{U}_{\mathcal{F}}$ der Wert $F([- \infty, \mu - \sigma(v) \mid \mu + \sigma, \omega])$ maximal wird, denn für konsistente Schätzer $\hat{\mu}$ und $\hat{\sigma}$ läßt sich die letztere Größe als Approximation für die Wahrscheinlichkeit für Werte außerhalb eines " $\bar{x} \pm s$ Bereiches" ansehen (vgl. auch die Theorie der sogenannten Toleranzbereiche etwa in Guttman 1970). Wegen der Gauß-Ungleichung (Karlin/Studden 1966, S. 483f) führt dies wieder auf die Gleichverteilung.
- (iii) Die Gleichverteilung hat unter allen Verteilungen aus $\mathcal{U}_{\mathcal{F}}$ den minimalen "tail-weight-index", wenn man diesen in Anlehnung an Rosenberger/Gasko (1983, S. 322f) definiert. Sie stellt insofern einen "Randpunkt" dar und kann als naheliegender Standard gebraucht werden.
- (iv) Über einem finiten Intervall ist die Gleichverteilung die Verteilung maximaler Entropie, vgl. hierzu 2.6.

2.5.3. Realitätsnähe:

Auch dieses Kriterium entzieht sich weitgehend der mathematischen Behandlung. Immerhin gibt Huber (1981, S. 91) ein Beispiel für einen sehr großen Datensatz, der gut durch eine 2%-kontaminierte Normalverteilung beschrieben wird. Für $\epsilon = 0.02$ liefert Tabelle 4.5.1. aus Huber (1981, S. 87) den Wert $k = 1.717$, der sich nicht viel vom "Konzentrationsellipsoidwert" $k = \sqrt{3} \approx 1.732$ unterscheidet. Als eine Art "benchmark" für robuste Schätzer kann Stiglers (1977) Studie dienen. Wir übernehmen hier Stiglers Vergleichskriterien; die Vor- und Nachteile dieser Methode sind in Stiglers Artikel und den Beiträgen der Diskussion ausführlich erörtert. Wir folgen zugleich Stiglers Vorschlag (1977, S. 1064) für die Überprüfung neuer Schätzer die "alten" s_j Werte zugrunde zu legen, da diese sich nicht erheblich verändern werden. Tabelle 2.2. (S. 61) zeigt die Resultate für die robusten Versionen von $(\bar{x}, s_n^2)$ und $(\bar{x}, s_{n-1}^2)$, wobei $s_n^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$ und $s_{n-1}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$ sei. Man beachte, daß s_{n-1}^2 keine "funktionale Interpretation", aber durchaus eine robustifizierte Version besitzt. Die Robustifizierung solcher Schätzer ist insbesondere in komplexen Modellen (z.B. Varianzkomponentenschätzung) von Interesse. Tabelle 2.3. (S. 61) faßt die Resultate von Tabelle 2.2. in den Indizes "relative error" (RE) und "relative rank" (RR) zusammen. Das Ergebnis ist sehr zufriedenstellend; das gilt auch im Vergleich mit der Nachfolgeuntersuchung von Balasooriya/Chan (1981).

2.5.4. Ergänzungen:

- (i) Da der oben betrachtete Schätzer ein M-Schätzer ist, lassen sich hierfür auch weitere Ergebnisse für M-Schätzer übernehmen bzw. entwickeln. Es sei diesbezüglich auf Boos/Serfling (1980) sowie insbesondere auf die Arbeiten von B.R. Clarke (1983, 1986) verwiesen, wo mit Methoden der "nonsmooth analysis" die Fréchet-Differenzierbarkeit von M-Funktionalen untersucht wird. Mit einem (fast sicheren) Invarianzprinzip beschäftigt sich Carroll (1978a).
- (ii) Statt die Funktionale

$$m^u(P) = \int x dP(x) \text{ und } v^u(P) = \int (x - m^u(P))^2 dP(x)$$

hätte man für eine fest gewählte Verteilung F und ein fest gewähltes $k > 0$ mit $E_F \psi^2(x) \leq 1$ (wobei ψ wie in 1.1.5.(8) sei)

Tabelle 2.2.: e_{ij} und r_{ij} für die Datensätze aus Stigler (1977)

Datensatz Nr.	robuste Version von $(\bar{x}, s_n^2)$		robuste Version von $(\bar{x}, s_{n-1}^2)$	
	e_{ij}	r_{ij}	e_{ij}	r_{ij}
1	0.95	3	0.93	3
2	0.96	5.5	0.96	5.5
3	0.41	2	0.33	2
4	0.90	3	0.89	3
5	0.05	1	0.14	1
6	0.93	3	0.91	3
7	1.00	6.5	1.00	6.5
8	0.99	5	0.99	5
9	0.92	10	0.92	10
10	0.98	4.5	0.98	4.5
11	0.98	3	0.98	3
12	0.98	3	0.97	3
13	1.02	9.5	1.02	9.5
14	1.02	6.5	1.02	6.5
15	1.01	8	1.01	8
16	1.02	10	1.02	10
17	0.88	2.5	0.86	2
18	0.87	3	0.86	2.5
19	0.95	6.5	0.96	7
20	0.90	2.5	0.90	2.5
21	0.87	5	0.86	4.5
22	0.92	3	0.91	3
23	0.94	3	0.94	3
24	1.01	10.5	1.01	10.5

Tabelle 2.3.: Relativer Fehler RE und Relativer Rang RR
für die Datensätze aus Stigler (1977)

	robuste Version von $(\bar{x}, s_n^2)$ $(\bar{x}, s_{n-1}^2)$	
RE(Small Samples)	0.886(0.24)	0.883(0.23)
RE(Large Samples)	0.935(0.06)	0.930(0.06)
RR(Small Samples)	4.9 (2.8)	4.9 (2.9)
RR(Large Samples)	5.4 (3.5)	5.3 (3.6)

auch die Funktionale

$$m^u(P) = \int x dP(x) \quad \text{und} \quad w^u(P) = \frac{1}{E_F \psi^2(x)} \cdot \int (x - m^u(P))^2 dP(x)$$

robustifizieren können. Dann ergibt sich ein Schätzer, der an der Stelle F konsistent ist (d.h. die Konsistenzkorrektur für die Scatterschätzung kann dort entfallen). Es ist jedoch nicht erforderlich, das "Robustifizieren" des Funktionals $v^u(P)$ und das "Konsistentmachen bei F " in dieser Weise zu verknüpfen. Wählt man aber etwa $F = \Phi = N(0,1)$, so ergibt sich der "Proposal 2" aus Huber (1964, S. 96f oder 1981, S. 137). Die Forderung $E_F \psi^2(x) = 1$ impliziert $k \geq 1$. Im Konzentrationsellipsoidfall wählt man zudem F als Gleichverteilung und daher $k = \sqrt{3}$.

- (iii) Die in Abschnitt 2.4. exemplarisch analysierte Robustifizierung läßt sich sowohl vertiefen wie auch verbreitern. Insbesondere die Bezugnahme auf das Momentenproblem legt den Gedanken nahe, durch Heranziehen z.B. des dritten Momentes asymmetrische Probleme zu behandeln. Dazu könnte etwa das Konzentrationsellipsoid bzgl. des Lokationswertes m unsymmetrisch gelegt oder aus zwei unterschiedlichen "Gleichverteilungen" zusammengesetzt werden. Aber auch für den Fall zweier Parameter ergeben sich weitere Anwendungsmöglichkeiten. Beispielsweise tritt gelegentlich der Fall auf, daß die Träger der Verteilungen sich verschieben. Als Beispiel kann etwa Example 6.10. aus Lehmann (1983, S. 452) dienen. Wir nehmen darüber hinaus an, daß die vorliegenden Daten nicht aus $E(\xi, b)$, sondern aus einer "nach oben ausgefranst" Exponentialverteilung stammen (d.h. der Schwanz der Verteilung ist durch Kontamination schwerer als exponentiell bzw. es gibt "Ausreißer nach oben"). Dieser Fall erscheint angesichts der Hauptanwendungen der Exponentialverteilung (z.B. Verteilungen von Lebensdauern und Wartezeiten) durchaus realistisch. Dann kann man aus den $\hat{\xi}$ und $\hat{b}$ (vgl. Lehmann 1983, S. 452) einen robustifizierten Schätzer für ξ und b gewinnen, indem man den (nicht kontaminierten) Schätzer $\hat{\xi}$ als linken Randpunkt des Konzentrationsellipsoides übernimmt und $\hat{b}$ in bekannter Weise durch Winsorisieren "von oben" robustifiziert.

2.6. Robustheit und Information

2.6.1. Situation: In diesem Abschnitt sind skizzenhaft einige grundsätzliche Überlegungen zur robusten Statistik und zur Wahl optimaler robuster Schätzer zusammengestellt. Viele hier benötigte Resultate sind in den Arbeiten von Fu (1973, 1975, 1982, 1985), Cheng/Fu (1986), Sievers (1978) und Kester/Kallenberg (1986) enthalten. Wir setzen bei Bezug auf Ergebnisse aus diesen Papieren stets implizit voraus, daß die dort jeweils angegebenen (technischen) Bedingungen erfüllt sind. Dabei betrachten wir hier nur den Spezialfall einer Familie $(P_{\theta})_{\theta \in \Theta}$ von W-Maßen auf $(\mathbb{R}, \mathcal{L})$ mit λ -Dichten $f(x, \theta) > 0$, die von einem eindimensionalen Parameter $\theta \in \Theta \subseteq \mathbb{R}$ abhängen. In Anlehnung an die Arbeiten von Bahadur kann dann die Güte einer Schätzerfolge (T_n) für einen Wert $g(\theta)$ durch die Abweichungsrate (inaccuracy) von (T_n)

$$e(\varepsilon, \theta, (T_n)) := -\liminf_{n \rightarrow \infty} \frac{1}{n} \cdot \log P_{\theta}(|T_n - g(\theta)| > \varepsilon) \quad (33)$$

gemessen werden, wobei ein guter Schätzer durch einen hohen Wert ausgezeichnet wird. Für konsistente Schätzerfolgen (T_n) gilt aber die Bahadur-Schranke

$$e(\varepsilon, \theta, (T_n)) \leq b(\varepsilon, \theta) := \inf_{\theta'} \{K(\theta', \theta) : |\theta' - \theta| < \varepsilon\}, \quad (34)$$

wobei

$$K_f(\theta', \theta) = \int \left\{ \log \frac{f(x, \theta')}{f(x, \theta)} \right\} \cdot f(x, \theta') dx \quad (35)$$

die Kullback-Leibler Information ist (vgl. Kester/Kallenberg 1986, S. 648f). Eine Schätzerfolge, die die Schranke (34) erreicht, heißt abweichungsratenoptimal. Es ist

$$K_f(\theta', \theta) = \int \log f(x, \theta') \cdot f(x, \theta') dx - \int \log f(x, \theta) \cdot f(x, \theta') dx. \quad (36)$$

Will man das Schätzproblem erschweren, so liegt es daher nahe, aus einer Menge etwa zur Wahl stehender Verteilungen diejenige mit

$$W_f(\theta') := \int \log f(x, \theta') \cdot f(x, \theta') dx = \text{Min!} \quad (37)$$

zu wählen. Dabei steht $W_f(\theta')$ für die "Wiener-Information", die von Wiener (1961, S. 76) als Konkurrent zur Fisher-Information vorgeschlagen wurde. Die Wiener-Information entspricht der negativen (Shannon-) Entropie, so daß der obige Vorschlag (37) auf die Anwendung des "Maximum-Entropie-Prinzips" hinausläuft.

2.6.2. Satz: Für zwei Verteilungsfamilien mit den Dichten $f(x, \theta)$ und $g(x, \theta)$ seien die Bedingungen 1 und 2 aus Fu (1982, S. 765) erfüllt. Dann gilt:

- (i) Falls es ein $\theta \in \mathbb{R}$ und ein $\varepsilon \in \mathbb{R}$ gibt derart, daß für $\tau \in \{\theta - \varepsilon, \theta + \varepsilon\}$ die Beziehung

$$\begin{aligned} W_f(\tau) &< W_g(\tau) + \int \log f(x, \theta) \cdot f(x, \tau) \, dx \\ &\quad - \int \log g(x, \theta) \cdot g(x, \tau) \, dx \end{aligned} \quad (38)$$

gilt, so folgt:

$$b_f(\varepsilon, \theta) < b_g(\varepsilon, \theta).$$

- (ii) Falls für die Fisher-Information an der Stelle θ die Beziehung $I_f(\theta) < I_g(\theta)$ gilt, so existiert eine Umgebung U von θ derart, daß für alle $0 \neq \varepsilon \in U$ folgt:

$$b_f(\varepsilon, \theta) < b_g(\varepsilon, \theta).$$

Beweis: Die Aussagen folgen unmittelbar wie im Beweis zu Theorem 3.1. von Fu (1982, S. 766).

2.6.3. Beispiel: Es sei $f(x, \theta) = \frac{1}{\sqrt{2\pi} \sigma_1} \cdot \exp \left[-\frac{(x-\theta)^2}{2\sigma_1^2} \right]$ und

$g(x, \theta) = \frac{1}{\sqrt{2\pi} \sigma_2} \cdot \exp \left[-\frac{(x-\theta)^2}{2\sigma_2^2} \right]$ mit $0 < \sigma_2 < \sigma_1$. Dann ist

$W_f(\theta) = -\frac{1}{2} \cdot \log(2\pi\sigma_1^2)$ und $K_f(\theta', \theta) = \frac{(\theta' - \theta)^2}{2\sigma_1^2}$. Somit ist für

alle θ und $\varepsilon > 0$ die Beziehung (38) erfüllt und $b_f(\varepsilon, \theta) < b_g(\varepsilon, \theta)$.

Wegen $I_f(\theta) = \frac{1}{\sigma_1^2} < \frac{1}{\sigma_2^2} = I_g(\theta)$ folgt speziell auch die Aussage von

(ii).

2.6.4. Bemerkung (ungünstigste Verteilungen): Satz 2.6.2. und Beispiel 2.6.3. beziehen sich auf ungünstigste Verteilungen im Sinne der Bahadur-Schranke. Allerdings wird die Bahadur-Schranke i.w. nur für Exponentialfamilien mit konvexem Parameterbereich tatsächlich erreicht; für das Lokations-Shift-Problem $f(x, \theta) = f(x - \theta)$ ist das i.a. nicht der Fall. In dieser Situation gelangt man zur Chernoff- oder Sievers-Schranke (vgl. Chernoff 1952, Sievers 1978, Kester/Kallenberg 1986)

$$e(\varepsilon, (T_n)) \leq C(-\varepsilon, \varepsilon) = -\log \inf_{0 < \alpha \leq 1} \int f^\alpha(x-\varepsilon) f^{1-\alpha}(x+\varepsilon) dx. \quad (39)$$

Diese Schranke gilt für alle translationsequivarianten (aber nicht notwendigerweise konsistenten) Schätzerfolgen. Im Gegensatz zur Bahadur-Schranke, die stets unter Benutzung der Kullback-Leibler Information definiert wird, wird der hinter der Sievers-Schranke stehende Informationsbegriff nicht explizit erwähnt. Die Parallelität der beiden Begriffe vom Standpunkt der Informationstheorie soll deshalb hier kurz dargestellt werden; wir verwenden dabei Begriffe, Resultate und Notationen aus Ferentinos/Papaioannou (1981). Es ist

$$\begin{aligned} C(-\varepsilon, \varepsilon) &= - \inf_{0 < \alpha < 1} \log \int f^\alpha(x) f^{1-\alpha}(x+2\varepsilon) dx \\ &= - \inf_{0 < \alpha < 1} (\alpha-1) \cdot I_x^R(f(x), f(x+2\varepsilon)), \end{aligned} \quad (40)$$

wobei $I_x^R(f_1, f_2)$ das durch

$$I_x^R(f_1, f_2) = \frac{1}{\alpha-1} \log \int f_1^\alpha f_2^{1-\alpha} d\mu$$

definierte nichtparametrische Rényi-Informationsmaß sei (Ferentinos/Papaioannou 1981, S. 199). Unter hinreichenden Regularitätsvoraussetzungen ergibt sich bei Benutzung der Ergebnisse aus Ferentinos/Papaioannou (1981, S. 203) für $\varepsilon = 0$ die Entwicklung

$$\begin{aligned} C(-\varepsilon, \varepsilon) &= \sup_{0 < \alpha < 1} - \log \int f^\alpha(x) f^{1-\alpha}(x+2\varepsilon) dx \\ &= \sup_{0 < \alpha < 1} \left\{ - \frac{\alpha(\alpha-1)}{2} (2\varepsilon)^2 \cdot I_f + O(2\varepsilon)^3 \right\} \\ &= \frac{1}{2} \varepsilon^2 \cdot I_f + O(2\varepsilon)^3. \end{aligned} \quad (41)$$

Dabei ist I_f die (in diesem Fall von ϑ unabhängige) Fisher-Information. Ausdruck (40) zeigt, daß das Schätzproblem wiederum dann besonders schwierig ist, wenn ein nichtparametrisches Informationsmaß klein wird; Ausdruck (41) zeigt wieder den Zusammenhang zur Fisher-Information (vgl. Fu 1982, S. 763). Die mittels Methode A aus Ferentinos/Papaioannou (1981, S. 200) aus den nichtparametrischen Informationsmaßen gewonnenen parametrischen Informationsmaße lassen sich ebenfalls für die Statistik nutzen. Beispielsweise entspricht die Beziehung $I_x^{KL}(\vartheta) = \frac{1}{2} \cdot I_x^F(\vartheta)$ (vgl. auch Aggarwal 1974, S. 115) i.w. der Aussage (1.2) in Fu (1975, S. 234). Als Richtlinien zur Auswahl von ungünstigsten Ver-

teilungen scheinen nach dem obigen demgegenüber Entropiemaße von größerem Interesse. Aus einem nichtparametrischen Maß kann man rein formal ein Entropiemaß erhalten, indem man die Information relativ zu der "Verteilung" mit der uneigentlichen Dichte $1_{\mathbb{R}}$ mißt.

Aus dem Kullback-Leibler-Maß $I_x^{KL}(f_1, f_2)$ erhält man so die Wiener-Information W_f ; aus dem Rényi-Maß $I_x^R(f_1, f_2)$ die Information

$$R_\alpha(f) := \frac{1}{\alpha-1} \log \int f^\alpha(x) dx, \quad \alpha > 0, \alpha \neq 1, \quad (42)$$

die wegen der Analogie zu Rényi (1961) als Rényi-Information der Ordnung α bezeichnet werde. Ohne näher darauf einzugehen, inwieweit $R_\alpha(f)$ - oder etwaige ähnlich hergeleitete Informationsbegriffe - plausible Eigenschaften von Informationsmaßen erfüllen, sei erwähnt, daß nach der Jensenschen Ungleichung $R_\alpha(f) \geq 0$ ist und außerdem $\lim_{\alpha \rightarrow 1} R_\alpha(f) = W_f$ gilt. Als allgemeines (d.h. in allen Situationen geeignetes und möglichst parameterunabhängiges) Maß für die Festlegung der ungünstigsten Verteilung erscheint daher die Wiener-Information ein geeigneter Kandidat. Satz 2.6.2. zeigt, wie dieses technisch in Erscheinung tritt.

2.6.5. Bemerkung (optimale Schätzer): Mit Blick auf optimale Schätzer im "large deviation" Sinne existieren die folgenden Resultate:

- (i) Der Maximum-Likelihood-Schätzer ist im eindimensionalen Fall genau dann optimal, wenn die zugrundeliegende Verteilung zu einer Exponentialfamilie gehört (Cheng/Fu 1986). Angenähert scheint dies auch zu gelten, falls die Krümmung der Verteilung klein ist (Fu 1982, S. 764).
- (ii) Im Lokations-Shift-Problem ist ein M-Schätzer (sogar finit) optimal (vgl. Kester/Kallenberg 1986, S. 655-660; die Voraussetzung " $\int (\log p) p dx < 0$ " in Theorem 3.4. auf S. 659 ist sogar entbehrlich, da ja $\int (\log p) p dx = -\int [(\log p)']^2 p dx$ und nicht $p \equiv \text{const.}$ ist).

2.6.6. Vorschlag (andere Sichtweise von Robustheit): Im Minimax-Ansatz zur robusten Statistik geht man stark vereinfacht so vor:

1. Man legt ein parametrisches Modell f und hierzu eine Umgebung $\mathcal{U}$ fest. (In der asymptotischen Minimaxtheorie zur Schätzung eines Lokationsparameters wird dabei $\mathcal{U}$ als kompakt in der vagen

Topologie der substochastischen Maße auf $\mathbb{R}$ unterstellt, vgl. Huber 1981, S. 76f.)

2. Man minimiert die Fisher-Information über $\mathcal{U}$ und wählt den Maximum-Likelihood-Schätzer zu dieser ungünstigsten Verteilung F_0 .

3. Angestrebt wird hierdurch die Minimierung der maximalen asymptotischen Varianz (für g aus $\mathcal{U}$) des Schätzers.

Der Influenzkurvenansatz arbeitet mit "infinitesimalen Umgebungen" (vgl. Hampel e.a. 1986, S. 40ff), verwendet aber im wesentlichen die gleiche Topologie (schwache Topologie) und das gleiche Zielkriterium (asymptotische Varianz).

Kritikpunkte an diesem Vorgehen sind u.a.:

- (i) Die "Unschärfevorstellung" wird nicht direkt in die Auswahl des parametrischen Modells f hineingenommen, sondern nachträglich durch Umgebungsbildung erfaßt.
- (ii) Die Topologie der schwachen Konvergenz ist nicht immer geeignet, "Nähe" von W -Maßen auszudrücken (vgl. Pfanzagl 1981, S. 16-18).

Das folgende Vorgehen ist in manchen Situationen daher vielleicht eher angezeigt:

- 1. Man wählt die ungünstigste Verteilung unter Berücksichtigung aller "sicheren" Information nach dem Maximum-Entropie-Prinzip (d.h. durch Minimierung der Wiener-Information).
- 2. Man wählt einen geeigneten Schätzer, etwa nach 2.6.5.(i) oder (ii).
- 3. Angestrebt wird hierdurch eine Optimalität im "large deviation" Sinne.

Wir illustrieren diesen Ansatz durch ein klassisches Beispiel:

2.6.7. Beispiel: Es sei der Lokationsparameter einer ein-dimensionalen Verteilung f zu schätzen; der Skalenparameter sei 1. Häufig wird mit Blick auf den Zentralen Grenzwertsatz angenommen, daß f eine Normalverteilung ist. Die Abschwächung, daß dies nur "approximativ" zu gelten braucht, hat Huber (1964, S. 74) zu einem Ausgangspunkt seiner Robustheitstheorie gemacht. Andererseits kann man die Annahme einer Normalverteilung auch anders begründen, nämlich als "uninformativste Verteilung" bei gegebenem ersten und zweiten Moment; d.h. als Verteilung mit minimaler Wiener-Information bei gegebenem $E(X)=g_1$ und $E(X^2)=g_2$. ("Gegeben" besagt

hier nicht, daß die Momente bekannt sind, sondern lediglich, daß sie für die beobachtete Zufallsvariable zugrunde lagen.) Eine Abschwächung dieser Annahme führt fast zwangsläufig zu dem obigen Vorschlag. Man betrachte zur Illustration die folgenden drei Fälle (vgl. Kagan/Linnik/Rao 1973, S. 408ff für die benötigten Resultate):

- (i) Man ist sicher, daß $E(X-\mu)^2$ oder damit gleichbedeutend $E(X)=g_1$ und $E(X^2)=g_2$ festliegen. Dann nimmt man eine Normalverteilung an und schätzt mit dem Maximum-Likelihood-Schätzer $\bar{X}$.
- (ii) Man ist nur sicher, daß $E|X-\mu| = g_1$ gegeben ist. Dann nimmt man eine Laplace-Verteilung an; als Maximum-Likelihood-Schätzer ergibt sich der Median.
- (iii) Man ist sicher, daß $E(\rho(X-\mu))$ mit ρ aus 1.1.(2) für bestimmtes k festliegt. Gegenüber (i) drückt diese Abschwächung gerade die "Unschärfe" aus, derentwegen Robustheit nötig ist. Dann ergibt sich als Verteilung mit minimaler Wiener-Information Hubers "ungünstigste Verteilung" f_0 (Huber 1981, S. 86). Der optimale Schätzer ist sowohl nach 2.6.5.(i) wie auch nach 2.6.5.(ii) Hubers Schätzer. Hubers Resultat bleibt also gültig. Die Interpretation und die Herleitung haben sich aber geändert.

2.6.8. Bemerkung: Den in 2.2. vorgeschlagenen Schätzer kann man also folgendermaßen motivieren:

- (i) Falls k bekannt ist, ist es der Schätzer nach 2.6.7.(iii).
- (ii) Falls k nicht bekannt ist, kann k aufgrund von 2.5. gewählt werden. Eine andere Sichtweise ist es, an die Verteilung überhaupt nur die Bedingung eines beschränkten Trägers zu stellen. Die Maximum-Entropie-Verteilung ist dann die Gleichverteilung. Hier ist bereits die Voraussetzung $f(x, \theta) > 0$ aus 2.6.1. verletzt, und ein brauchbarer Schätzer kann mit Hilfe von 2.6.5. nicht bestimmt werden. (Der MLE $\frac{1}{2}(x_{(1)} + x_{(n)})$ ist zumindest für einige Verteilungen aus der entsprechenden "Entropieumgebung" - d.h. aus der Menge der Verteilungen, die mindestens den formulierten Bedingungen genügen - offensichtlich ungeeignet.) Man greift deshalb auf eine weitere sehr allgemeine Schätzmethode - die Momentenmethode - zurück und robustifiziert diesen Schätzer.

2.6.9. Notiz (über Realitätsnähe): Durch Verschmelzung ("merger") von Verteilungen lassen sich Verteilungen gewinnen, die zumindest für die Datenanalyse interessant sind (vgl. Olbricht 1987). Hubers ungünstigste Verteilung f_0 ergibt sich in diesem Ansatz als linearer φ -merger zweier Exponentialverteilungen, d.h. als nahe-
liegende Verteilung für ein "Zwei-Tendenzen-Modell" mit Minimal-
annahmen. Da die derart gewonnenen Verteilungen nahe an der hyper-
bolischen Verteilung liegen, spricht die für letztere gefundene
empirische Evidenz (Barndorff-Nielsen 1977 und 1979) auch für den
hier vorgeschlagenen Maximum-Entropie-Ansatz.

2.6.10. "Philosophische Notizen":

- (i) In diesem Abschnitt wurde desöfteren auf "Prinzipien" und heuristische Überlegungen zurückgegriffen. Dies soll mathe-
tische Aussagen aber nicht ersetzen, sondern ergänzen. Man
befindet sich an diesen Stellen im Vorfeld mathematischer
Untersuchungen, wo man - wie es scheint - "aus Prinzip"
nicht auf Prinzipien verzichten kann (vgl. auch Müller
1984).
- (ii) Von zentraler Bedeutung für den hier vertretenen Ansatz ist
die Wiener-Information, die wegen ihrer Identität mit der
negativen (Shannon-) Entropie die bekannten informations-
theoretischen, mathematischen und physikalischen Interpre-
tationen und Anwendungen besitzt. Die beiden ersteren sind
weitgehend dual zu den entsprechenden Konzepten für die
Fisher-Information (ungünstigste Verteilung, Minimax-Eigen-
schaft im Effizienz bzw. "large deviation" Sinne). Der phy-
sikalischen Interpretation nach ist es nicht erstaunlich,
daß die sich ergebenden ungünstigsten Verteilungen "reali-
tätsnah" sind (vgl. Huber 1964, S. 75 und auch van Campen-
hout/Cover 1981).
- (iii) Der Ansatz berührt allgemeinere Konzepte in Informations-
theorie, Mathematik und Physik. Beispielsweise entspricht
die Aufspaltung (36) dem Zusammenhang von "error", "entropy"
und "inaccuracy" nach Kerridge (1961, insbesondere S. 193).
Mittels allgemeinerer Entropiebegriffe (Lanford 1973) und
"large deviation" Theorien (Varadhan 1984, Donsker/Varadhan
1975, 1976) könnte man Robustheit auch für kompliziertere
Strukturen (Markov Prozesse) nach den gleichen Prinzipien

untersuchen (vgl. zum Hintergrund Ellis 1985). Aus angewandter Sicht erscheint dies derzeit nicht spruchreif, da selbst der eindimensionale Lokationsfall in dieser Hinsicht nicht beherrscht wird.

- (iv) Bemerkenswert ist schließlich die enge Verbindung von Maximum-Entropie-Problemen und dem Momentenproblem (vgl. auch Landau 1987). Man wird i.a. die Maximum-Entropie-Verteilung dadurch festlegen, daß man möglichst weitreichende Momentenbedingungen vorgibt. Es spricht für die Geschlossenheit des Ansatzes, daß dann sowohl die "Ausgangsumgebungen" wie auch die resultierenden Äquivalenzklassen über Momente charakterisiert werden. Benutzt man noch die Momentenmethode als Schätzprinzip, so kann man kurz formulieren:
Geschätzt werden die Größen, die auch sicher gegeben sind; die "Unschärfe" wird durch eine Äquivalenzrelation ausgedrückt.

Wir geben zum Abschluß drei Zitate an, die - obschon in anderen Zusammenhängen stehend - auch die obigen Ideen beleuchten. Zur Bedeutung von Momenten (im Unterschied zu weniger durchsichtigen Überlegungen) bemerkte E. S. Pearson (1962, S. 394):

I sometimes think that Statistics is becoming far too mathematical, and that it is a relief to turn to the many simple, unsolved problems which can be discussed in terms only of means and standard deviations.

Wie ein Programm zur robusten Statistik liest sich Jaynes (1957, S. 623):

It is now evident how to solve our problem; in making inferences on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known. This is the only unbiased assignment we can make; to use any other would amount to arbitrary assumption of information which by hypothesis we do not have.

Schließlich geht es nach Weyl (1974, S. 5) bei Äquivalenzklassenbildungen um folgendes:

Dieses Prinzip wird namentlich dann immer anzuwenden

sein, wenn uns an den Objekten $a, b, \dots$ nur diejenigen Eigenschaften interessieren, welche gegenüber der Beziehung $\sim$ invariant sind. Seine Anwendung hat den Erfolg, daß eine schwerfällige Ausdrucksweise durch eine kürzere ersetzt wird, die dem herrschenden Interessestandpunkt der Untersuchung dadurch Rechnung trägt, daß sie von selbst alles im Sinne dieses Standpunktes Unwesentliche an den untersuchten Objekten abstreift.

Wo man am meisten drauf erpicht,
 Grad das bekommt man meistens nicht.
 (W. Busch)

3. LOKATIONS- UND SKALENSCHÄTZUNG IM MEHRDIMENSIONALEN FALL

In diesem Kapitel werden die Konzepte aus Kapitel 1 auf die Lokations- und Skalenschätzung im mehrdimensionalen Fall angewendet. Die statistische Theorie multivariater M-Schätzer wurde vor allem in den Arbeiten von Maronna (1976), Huber (1977) und Schönholzer (1979) ausgearbeitet. Wir werden uns hier vorwiegend auf die fast völlig offene Frage der Eindeutigkeit der Schätzungen konzentrieren.

3.1. Allgemeines

3.1.1. Situation: Es liege eine Lokations- und Skalenfamilie $F_{\mu, \Sigma}(x) = F(\sqrt{\Sigma^{-1}} \cdot (x - \mu))$ vor, wobei F eine Verteilungsfunktion über $\mathbb{R}^p$, $\mu \in \mathbb{R}^p$, Σ eine positiv definite $(p \times p)$ -Matrix und $\sqrt{\Sigma^{-1}}$ die (eindeutig bestimmte) symmetrische positiv definite Wurzel aus Σ^{-1} sei. Um robuste Schätzer für μ und Σ zu erhalten, robustifizieren wir nach den obigen Methoden die Standardschätzer

$$m^u(P) = \int x dP(x) \quad \text{und} \quad V^u(P) = \int (x - m^u(P))(x - m^u(P))^T dP(x)$$

für W -Maße P auf $\mathbb{R}^p$. Als Norm wird die Euklidische Norm $\|\cdot\|_2$, als Ellipsoide die kanonischen Ellipsoide $\{x \in \mathbb{R}^p \mid \|\sqrt{\Sigma^{-1}}(x - m)\|_2^2 \leq k^2\}$ (Lokationsfunktion = Identität, Scatterfunktion = symmetrische positiv definite Wurzel, "tuning constant" = k^2) und als Projektion die radiale Projektion auf den Mittelpunkt hin (vgl. 1.2.2.(6)) benutzt. Für das Konzentrationsellipsoid ist $k = \sqrt{p+2}$ zu wählen. Es ist $\gamma = (m, V) \in \Gamma = \mathbb{R}^p \times \{V \mid V \text{ ist p.d. } (p \times p)\text{-Matrix}\}$ und für ein W -Maß P auf $(\mathbb{R}^p, \mathcal{L}^p)$ gilt:

$$d(p, e(\gamma), P) =: d((m, V), P) = \left[\int \psi(x; m, V) dP(x) \right. \\ \left. \int \left[\psi(x; m, V) - m \right] \cdot \left[\psi(x; m, V) - m \right]^T dP(x) \right], \quad (1)$$

wobei die Funktion $\psi(x; m, V)$ durch

$$\psi(x; m, V) = \begin{cases} x & , \| \sqrt{\Sigma^{-1}}(x-m) \|_2^2 \leq k^2 \\ \frac{k}{\| \sqrt{\Sigma^{-1}}(x-m) \|_2} (x-m) + m & , \| \sqrt{\Sigma^{-1}}(x-m) \|_2^2 > k^2 \end{cases} \quad (2)$$

gegeben sei. Gesucht sind bei gegebenem P die Fixpunkte der Abbildung

$$d: (m, V) \longmapsto d((m, V), P). \quad (3)$$

3.1.2. Notiz: Die Gleichung

$$d((m_0, V_0), P) = (m_0, V_0)$$

läßt sich nach (1) und (2) auch in der Form

$$\begin{aligned} E_P u_1(\| \sqrt{V_0^{-1}}(x-m_0) \|_2)(x-m_0) &= 0 \\ E_P u_2(\| \sqrt{V_0^{-1}}(x-m_0) \|_2^2)(x-m_0)(x-m_0)^T &= V_0 \end{aligned} \quad (4)$$

mit $u_1(s) := \min\left[1, \frac{k}{s}\right]$ und $u_2(s) := \min\left[1, \frac{k^2}{s}\right]$ für $s \geq 0$ schreiben. Daher lassen sich die Resultate aus Maronna (1976) direkt nutzen. Es ist in der Notation von Maronna (1976) $\psi_1(s) = \min(s, k)$, $\psi_2(s) = \min(s, k^2)$, $K_1 = k$, $K_2 = k^2$. Die dortigen Bedingungen (A), (B) und (C) sind also erfüllt; Bedingung (D) ist gleichbedeutend mit $k^2 > p$ (und sicher für $k = \sqrt{p+2}$ erfüllt). Wir notieren zwei wichtige Ergebnisse:

3.1.3. Satz (Existenz): Es existiere ein $a > 0$ derart, daß für jede Hyperebene H die Beziehung $P(H) \leq 1 - \frac{p}{k^2} - a$ gilt. Dann existiert ein Fixpunkt (m_0, V_0) von (3). Ferner ist $m_0 \in \text{co}(\text{supp}(P))$.

Beweis: Maronna (1976, Theorem 2, S. 57) bzw. Schönholzer (1979, Satz 3, S. 31ff) und 3.1.2.

3.1.4. Satz (Konsistenz): Es existiere ein $a > 0$ derart, daß für jede Hyperebene H die Beziehung $P(H) \leq 1 - \frac{p}{k^2} - a$ gilt. Ferner besitze P einen eindeutig bestimmten Fixpunkt (m_0, V_0) von (3). Bezeichnet dann P^n die empirische Verteilung einer Stichprobe vom Umfang n und (m_n, V_n) einen Fixpunkt von (3) für P^n , so gilt mit

$$\lim_{n \rightarrow \infty} (m_n, V_n) = (m_0, V_0)$$

Beweis: Maronna (1976, Theorem 5, S. 59) und 3.1.2.

3.1.5. Notiz: Maronna (1976, S. 61) gibt in Theorem 6 auch Bedingungen für die asymptotische Normalität, die aber wegen Differenzierbarkeitsvoraussetzungen an die Funktionen ψ_1 und ψ_2 nicht direkt anwendbar sind. Man kann jedoch wiederum direkt auf Hubers (1981, S. 132f) Bedingungen zurückgreifen. Weitere Aussagen zur asymptotischen Verteilung - insbesondere über ein LIL - findet man in Carroll (1978b).

Es sei Q eine Verteilung auf $(\mathbb{R}^p, \mathcal{L}^p)$ mit einem eindeutig bestimmten Fixpunkt (m_0, V_0) von (3). (Einige Bedingungen hierfür gibt etwa Theorem 3 aus Maronna (1976, S. 58). Trivialerweise gilt dies für die Gleichverteilung über dem zu (m_0, V_0) gehörenden Konzentrationsellipsoid.) Ähnlich wie in 2.4. betrachte man zur Untersuchung der Eindeutigkeit des Fixpunktes für ein weiteres W -Maß W auf $(\mathbb{R}^p, \mathcal{L}^p)$ die Abbildung

$$H(\gamma, t) = d(p, e(\gamma), t \cdot W + (1-t) \cdot Q) - \gamma \quad (5)$$

mit $\gamma = (m, V)$.

3.1.6. Bemerkung (Reduktion durch Diagonalisierung): Es ist hinreichend, den Fall $V = \text{diag}(v_{11})_{i=1, \dots, p}$ zu untersuchen, wobei $\text{diag}(v_{11})_{i=1, \dots, p}$ die $(p \times p)$ -Diagonalmatrix mit den (hier positiven) Diagonalelementen v_{11} bezeichne. Existieren nämlich zu einem P zwei Fixpunkte (m_1, V_1) und (m_2, V_2) von (3), so kann man nach dem Satz über die simultane Diagonalisierung von p.d. Matrizen (Searle 1982, Theorem 2, S. 313) eine nichtsinguläre Matrix R finden mit $RV_1R^T = I$ und $RV_2R^T = D$, wobei D eine Diagonalmatrix ist. Ist P die Verteilung einer Zufallsvariablen X , so besitzt die Verteilung P' von $Y := RX - Rm_1$ die Fixpunkte $(0, I)$ und $(R(m_2 - m_1), D)$. Betrachtet man in dieser Weise das Problem nur in den $2p$ Parametern $\gamma = ((m_1, \dots, m_p)^T, (v_{11}, \dots, v_{pp})^T)^T$, so sprechen wir auch vom "reduzierten Problem". Es entsteht aus dem "vollen Problem" dadurch, daß nur die Mittelwerte und Varianzen geschätzt werden und die Kovarianzen (für das Konzentrationsellipsoid) auf 0 gesetzt werden. Nach den obigen Überlegungen ist das "reduzierte Problem" in

Bezug auf die Eindeutigkeit hinreichend für das "volle Problem".
(Es ist sogar schärfer, da die Kovarianzbedingungen verletzt sein können.) Zur Vereinfachung der Notation schreiben wir auch im "reduzierten Problem" $(0, I)$ statt etwa $((0, 0)^T, (1, 1)^T)$.

3.1.7. Bemerkung (Reduktion durch Equivarianz): Im "reduzierten Problem" sei $(\gamma_0, t_0) = ((m_0, V_0), t_0) = ((m_0, \text{diag}(v_{11}^0)_{i=1, \dots, p}), t_0)$ eine beliebige feste Nullstelle von H . Mit der unten in (6) definierten Abbildung $H'(\gamma, t)$ gilt dann

$$\pi_{\gamma}(\partial H(\gamma_0, t_0)) = \{D_2 \cdot D \cdot D_1 \mid D \in \pi_{\gamma}(\partial H'((0, I), t_0))\}.$$

Dabei ist

$$D_1 = \text{diag}(v_{11}^0)_{i=1, \dots, 2p} \text{ mit } v_{11} = \begin{cases} \frac{1}{\sqrt{v_{11}^0}} & , \text{ falls } i=1, \dots, p \\ \frac{1}{v_{11}^0} & , \text{ falls } i=p+1, \dots, 2p \end{cases}$$

$$D_2 = \text{diag}(v_{11}^0)_{i=1, \dots, 2p} \text{ mit } v_{11} = \begin{cases} \sqrt{v_{11}^0} & , \text{ falls } i=1, \dots, p \\ v_{11}^0 & , \text{ falls } i=p+1, \dots, 2p \end{cases}$$

Beweis: Analog wie in 2.4.2.1. seien W' und Q' durch

$$W'(B) := W(\{\sqrt{V_0^{-1}}(x - m_0) \mid x \in B\}) \quad \forall B \in \mathcal{B}^p$$

$$Q'(B) := Q(\{\sqrt{V_0^{-1}}(x - m_0) \mid x \in B\}) \quad \forall B \in \mathcal{B}^p$$

sowie die Transformationen

$$T'(m, V) := \begin{bmatrix} \sqrt{V_0^{-1}} \cdot (m - m_0) \\ V_0^{-1} \cdot V \end{bmatrix} \quad \text{und} \quad T''(m, V) := \begin{bmatrix} \sqrt{V_0} \cdot m + m_0 \\ V_0 \cdot V \end{bmatrix}$$

gegeben. Dann ist $((0, I), t_0)$ eine Nullstelle von

$$H'(\gamma, t) := d(p, e(\gamma), t \cdot W' + (1-t) \cdot Q') - \gamma, \quad (6)$$

und es gilt

$$H(\gamma, t) = T''(H'(T'(\gamma), t)) - \begin{bmatrix} m_0 \\ 0_p \end{bmatrix},$$

woraus die Behauptung folgt.

Wir führen noch einige Vereinfachungen ein:

3.1.8. Vereinbarungen: Sofern nicht ausdrücklich anders erwähnt, sei für die Abschnitte 3.2. und 3.3. vereinbart:

- (i) Die betrachteten Verteilungen über $\mathbb{R}^p$ besitzen eine λ^p -f.ü. stetige λ^p -Dichte. Damit ist die Differenzierbarkeit von H nach γ gesichert. Den Fall diskreter Maße kann man im Prinzip durch nonsmooth Analysis einbeziehen, worauf hier aber verzichtet wird (vgl. auch 1.3.3.3.(ii)).
- (ii) Die Voraussetzungen (i), (ii) und (iii) aus 1.3.2.5. seien jeweils erfüllt, d.h. die Untersuchung kann sich ganz auf die Regularität der kritischen Matrix

$$H_{\gamma}((0, I), t_0) \quad (7)$$

konzentrieren.

- (iii) Wir betrachten zur notationsmäßigen Vereinfachung nur den Fall $k = \sqrt{p+2}$, d.h. den des Konzentrationsellipsoids. Dies ist auch der Grund, weshalb in 3.1.6. anders als in 2.4. $(0, I)$ als Standardpunkt gewählt wurde. (In 2.4. war $\left[0, \frac{1}{k^2}\right]$ betrachtet worden, um für jedes k die Beziehung $k\sqrt{v_0} = 1$ zu erreichen.)

3.2. Teilaussagen im zweidimensionalen Fall

In diesem Abschnitt sei stets $p = 2$. Es ist dann günstig, Polarkoordinaten zu benutzen:

3.2.1. Satz: Es sei (γ_0, t_0) mit $\gamma_0 = (0, I) = ((0, 0)^T, (1, 1)^T)^T$ eine Nullstelle von H aus (5) für das "reduzierte Problem". Mit $P := t_0 \cdot W + (1 - t_0) \cdot Q$, $A := \{(r, \varphi) \mid r \leq 2, 0 \leq \varphi \leq 2\pi\}$, $B := \{(r, \varphi) \mid r > 2, 0 \leq \varphi \leq 2\pi\}$ und $q := P(A)$ gilt dann:

- (i) Die Gleichung $H(\gamma_0, t_0) = 0$ entspricht dem Gleichungssystem

$$\int_A r \cos \varphi \, dP(r, \varphi) + \int_B 2 \cos \varphi \, dP(r, \varphi) = 0 \quad (8.a)$$

$$\int_A r \sin \varphi \, dP(r, \varphi) + \int_B 2 \sin \varphi \, dP(r, \varphi) = 0 \quad (8.b)$$

$$\int_A r^2 \cos^2 \varphi \, dP(r, \varphi) + \int_B 4 \cos^2 \varphi \, dP(r, \varphi) = 1 \quad (8.c)$$

$$\int_A r^2 \sin^2 \varphi \, dP(r, \varphi) + \int_B 4 \sin^2 \varphi \, dP(r, \varphi) = 1 \quad (8.d)$$

(ii) Die kritische Matrix (7) ist

$$H_{\gamma}(\gamma_0, t_0) = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \quad (9)$$

mit

$$H_{11} = \begin{bmatrix} -q - \int_B 2 \cdot \frac{\sin^2 \varphi}{r} \, dP(r, \varphi) & \int_B 2 \cdot \frac{\cos \varphi \sin \varphi}{r} \, dP(r, \varphi) \\ \int_B 2 \cdot \frac{\cos \varphi \sin \varphi}{r} \, dP(r, \varphi) & -q - \int_B 2 \cdot \frac{\cos^2 \varphi}{r} \, dP(r, \varphi) \end{bmatrix}$$

$$H_{12} = \begin{bmatrix} \int_B \cos^3 \varphi \, dP(r, \varphi) & \int_B \cos \varphi \sin^2 \varphi \, dP(r, \varphi) \\ \int_B \cos^2 \varphi \sin \varphi \, dP(r, \varphi) & \int_B \sin^3 \varphi \, dP(r, \varphi) \end{bmatrix}$$

$$H_{21} = \begin{bmatrix} -\int_A 2r \cos \varphi \, dP(r, \varphi) - 8 \cdot I_1 & 8 \cdot I_2 \\ 8 \cdot I_1 & -\int_A 2r \sin \varphi \, dP(r, \varphi) - 8 \cdot I_2 \end{bmatrix}$$

$$\text{mit } I_1 := \int_B \frac{\cos \varphi \sin^2 \varphi}{r} \, dP(r, \varphi)$$

$$\text{und } I_2 := \int_B \frac{\cos^2 \varphi \sin \varphi}{r} \, dP(r, \varphi)$$

$$H_{22} = \begin{bmatrix} -1 + 4 \int_B \cos^4 \varphi \, dP(r, \varphi) & 4 \int_B \cos^2 \varphi \sin^2 \varphi \, dP(r, \varphi) \\ 4 \int_B \cos^2 \varphi \sin^2 \varphi \, dP(r, \varphi) & -1 + 4 \int_B \sin^4 \varphi \, dP(r, \varphi) \end{bmatrix}$$

Beweis: Die Aussagen ergeben sich durch Verifizieren; bei (ii) ist

unter dem Integralzeichen zu differenzieren.

3.2.2. Bemerkung: In der Situation von 3.2.1. gilt:

- (i) $\det(-H_{11}) > 0$
- (ii) $\det(-H_{22}) > 0$
- (iii) Falls P punktsymmetrisch bezüglich des Nullpunktes ist
(d.h. $P(X \in A) = P(-X \in A) \quad \forall A \in \mathcal{L}^p$), so ist
 $\det(-H) = \det(-H_{11}) \cdot \det(-H_{22}) > 0$.

Beweis: (i) Durch Addition von (8.c) und (8.d) folgt $q > \frac{1}{2}$. Wegen

$$\int_2 \frac{\cos \varphi \cdot \sin \varphi}{r} dP(r, \varphi) \leq 1 - q < \frac{1}{2} < q \text{ ist } (-H_{11}) \text{ diagonaldominant.}$$

(ii) Nach (8.c) und (8.d) ist

$$1 - 4 \int_B \cos^2 \varphi dP(r, \varphi) = \int_A r^2 \cos^2 \varphi dP(r, \varphi) > 0$$

$$1 - 4 \int_B \sin^2 \varphi dP(r, \varphi) = \int_A r^2 \sin^2 \varphi dP(r, \varphi) > 0.$$

Daher ist $(-H_{22})$ diagonaldominant.

(iii) ist klar.

Die in 3.2.2.(i) und (ii) behandelten Fälle entsprechen im wesentlichen den Situationen, daß entweder m oder V bereits festliegen. Derartige Fälle lassen sich leichter direkt behandeln (Huber 1981, S. 215ff); der Homotopieansatz würde aber auch zum Ziel führen. Ganz allgemein läßt sich nur wenig über (9) sagen:

3.2.3. Satz (Diagonaldominanzfall): Falls in der Situation aus

3.2.1. $q > \frac{9}{10}$, so ist $\det(-H_\gamma(\gamma_0, t_0)) > 0$.

Beweis: Mit Hilfe der Abschätzung $\cos \varphi \cdot \sin \varphi = \frac{1}{2} \sin(2\varphi) \leq \frac{1}{2}$ ergibt sich in diesem Falle Diagonaldominanz.

3.2.4. Satz ("Orbitalfall"): Es liege die Situation von 3.2.1. vor. Es sei zusätzlich

$$\int_A r^2 \sin \varphi \cos \varphi dP(r, \varphi) + \int_B 4 \sin \varphi \cos \varphi dP(r, \varphi) = 0 \quad (10)$$

(d.h. auch die Kovarianzbedingung ist erfüllt; der Fixpunkt entspricht also einem echten Fixpunkt im Sinne des "vollen Prob-

lems"). Ferner besitze P eine Dichte der Form $h(r, \varphi) = f(r) \cdot g(\varphi)$
 $\forall r \geq 0 \quad \forall \varphi \in [0, 2\pi]$. Dann ist $\det (-H_{\gamma}(\gamma_0, t_0)) > 0$.

Beweis: Aus (8.a) folgt $\int_0^{2\pi} \cos \varphi \, g(\varphi) d\varphi = 0$. Aus (8.c) und (8.d)

erhält man $\int_0^{2\pi} \cos^2 \varphi \, g(\varphi) d\varphi = \int_0^{2\pi} \sin^2 \varphi \, g(\varphi) d\varphi = \frac{1}{2}$. Schließlich folgt

aus (10): $\int_0^{2\pi} \cos \varphi \sin \varphi \, g(\varphi) d\varphi = 0$. Damit ergibt sich für die kriti-

sche Matrix (9) nach

- Multiplikation der 3. und 4. Zeile mit $\frac{1}{4}$
- Addition der 3. zur 4. Spalte
- Addition der 3. zur 4. Zeile

die folgende Matrix

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

wobei

$$A_{11} = \left[q + \int_{r>2} \frac{1}{r} f(r) dr \right] \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$A_{12} = \begin{pmatrix} (1-q) \cdot \int_0^{2\pi} \cos \varphi \sin^2 \varphi \, g(\varphi) d\varphi & 0 \\ 0 & - (1-q) \int_0^{2\pi} \cos^2 \varphi \sin \varphi \, g(\varphi) d\varphi \end{pmatrix}$$

$$A_{21} = \left[\int_{r>2} \frac{2}{r} f(r) dr \right] \cdot \begin{pmatrix} \int_0^{2\pi} \cos \varphi \sin^2 \varphi \, g(\varphi) d\varphi & - \int_0^{2\pi} \cos^2 \varphi \sin \varphi \, g(\varphi) d\varphi \\ 0 & 0 \end{pmatrix}$$

$$A_{22} = \begin{pmatrix} \frac{1}{4} - \frac{1}{2}(1-q) + (1-q) \cdot \int_0^{2\pi} \cos^2 \varphi \sin^2 \varphi \, g(\varphi) d\varphi & \frac{1}{4} - \frac{1}{2}(1-q) \\ \frac{1}{4} - \frac{1}{2}(1-q) & \frac{1}{2} - (1-q) + 2(1-q) \cdot \int_0^{2\pi} \cos^2 \varphi \sin^2 \varphi \, g(\varphi) d\varphi \end{pmatrix}$$

Nach der Determinantenformel für partitionierte Matrizen (z.B. Mardia/Kent/Bibby 1979, (A.2.3j), S. 457) gilt

$$\det A = \det(A_{11}) \cdot \det(A_{22} - A_{21} \cdot A_{11}^{-1} \cdot A_{12}).$$

Wegen der Jensenschen Ungleichung ist

$$\begin{aligned} & \frac{1}{q + \int_{r>2} \frac{1}{r} f(r) dr} \cdot (1-q) \left\{ \int_{r>2} \frac{2}{r} f(r) dr \right\} \cdot \left[\left[\int_0^{2\pi} \cos \varphi \sin^2 \varphi g(\varphi) d\varphi \right]^2 \right. \\ & \quad \left. + \left[\int_0^{2\pi} \cos^2 \varphi \sin \varphi g(\varphi) d\varphi \right]^2 \right] \\ & \leq (1-q) \int_0^{2\pi} \cos^2 \varphi \sin^2 \varphi g(\varphi) d\varphi. \end{aligned}$$

Somit folgt die Behauptung.

3.2.5. Notiz: Mittels der in 3.2.2. - 3.2.4. genannten Kriterien für die Regularität der Determinante im "reduzierten Problem" (und gegebenenfalls ähnlicher Überlegungen für das "volle Problem") kann man Voraussetzungen formulieren, unter denen die Eindeutigkeit gilt. Aus statistischer Sicht ist es allerdings von größerem Interesse, statt Teillösungen für das vollständige Problem vollständige Lösungen für Teilprobleme zu finden. Dies gilt umso mehr, als die reine mehrdimensionale Lokations- und Scatterschätzung ein eher seltenes statistisches Problem ist. In den meisten Fällen liegt weitere Information vor oder werden vertiefende multivariate Analyseverfahren angestrebt. Der Hinweis, daß die Lokations- und Scatterschätzung als Ausgangspunkt dienen und mittelbar zur Robustifizierung nahezu sämtlicher multivariater Verfahren benutzt werden kann, ist zwar richtig, baut aber eine unnötig hohe Hürde auf. Daß sich trotzdem fast alle Untersuchungen auf diesen Fall konzentrieren (- allerdings mit Blick auf die Eindeutigkeit nahezu ergebnislos geblieben sind -), läßt sich wohl auch damit erklären, daß die robusten Schätzer nicht flexibel genug für andere Modelle "angepaßt" werden konnten. Bei dem in Kapitel 1 beschriebenen Schätzprinzip ist dies hingegen leicht möglich. (Ein Beispiel ist bereits das "reduzierte Problem".)

Im folgenden setzen wir die Vereinbarung 3.1.8.(ii) nicht mehr voraus und betrachten einen sehr speziellen Fall, der in 4.3.10. nochmals kurz angesprochen werden wird.

3.2.6. Situation: Auf $(\mathbb{R}^2, \mathcal{L}^2)$ betrachte man das Lokations-Scatter-Modell $F_{\mu, \Sigma}(x) = F(\sqrt{\Sigma}^{-1}(x-\mu))$, wobei

$$\mu = \begin{bmatrix} \mu_0 \\ 0 \end{bmatrix} \quad \text{und} \quad \Sigma = \begin{bmatrix} a & 0 \\ 0 & \sigma_0^2 \end{bmatrix} \quad (11)$$

mit $\mu_0 \in \mathbb{R}$ und $\sigma_0^2 \in]0, \infty[$ sei. a sei eine bekannte positive Konstante, die o.B.d.A. 1 gesetzt sei. Als Schätzer für μ_0 und σ_0^2 werden die Standardschätzer

$$m(P) = \int x_1 dP(x) \quad \text{und} \quad v(P) = \int x_2^2 dP(x) \quad (12)$$

(x_i := i-te Komponente von x) - als Funktionale auf dem Raum der W -Maße über $(\mathbb{R}^2, \mathcal{L}^2)$ mit den entsprechenden Momenten - verwendet. Bei der Robustifizierung von (12) bildet man die Konzentrations-ellipsoide gemäß (11). P_1 sei ein W -Maß auf $(\mathbb{R}^2, \mathcal{L}^2)$ mit einer λ^2 -f.s. stetigen λ^2 -Dichte und $\int (x_1 - m(P_1))^2 dP_1(x) \leq 1$.

3.2.7. Satz: In der Situation von 3.2.6. gibt es für P_1 genau einen Fixpunkt (m', v') von d aus (3).

Beweis: Wir bezeichnen mit Q die Gleichverteilung auf der Kreisscheibe um $m(P_1)$ mit Radius 2, setzen $W = P_1$ sowie $P_t := t \cdot W + (1-t) \cdot Q$ für $t \in [0, 1]$ und wenden Satz 1.3.2.5. auf die Abbildung $H(\gamma, t)$ aus (5) an. Die Voraussetzungen (i) und (iii) sind offensichtlich erfüllt. Zum Beweis von (ii) beachte man, daß für jeden Fixpunkt (m_0, v_0) m_0 in einem festen beschränkten Bereich liegen muß. Es bleibt daher nur die Regularität der kritischen Matrix $H_\gamma((m_0, v_0), t_0)$ in einem beliebigen Fixpunkt (m_0, v_0) eines Maßes P_{t_0} zu zeigen. O.B.d.A. kann man sich wieder auf den Fall $(m_0, v_0) = (0, 1)$ beschränken. Die Situation entspricht dann dem in 3.2.1. betrachteten Fall, wenn man dort $m_2 \equiv 0$ und $v_2 \equiv 1$ als bekannt ansieht. Bei Benutzung von Polarkoordinaten und der Bezeichnung P für das transformierte (aber nicht näher spezifizierte) W -Maß entspricht das Gleichungssystem $H(\gamma_0, t_0) = 0$ den Gleichungen (8.a) und (8.d), und für die kritische Matrix erhält man die vier "Eckpunkte" der Matrix (9), d.h.

$$H_{\gamma}(\gamma_0, t_0) = \begin{pmatrix} -q - 2 \cdot \int_B \frac{\sin^2 \varphi}{r} dP(r, \varphi) & \int_B \cos \varphi \sin^2 \varphi dP(r, \varphi) \\ 8 \cdot \int_B \frac{\cos \varphi \sin^2 \varphi}{r} dP(r, \varphi) & -1 + 4 \int_B \sin^4 \varphi dP(r, \varphi) \end{pmatrix} \quad (13)$$

mit A, B und q wie in 3.2.1. Zum Nachweis der Regularität von (13) genügt es, die positive Definitheit der Matrix

$$\begin{pmatrix} q + 2 \cdot \int_B \frac{\sin^2 \varphi}{r} dP(r, \varphi) & - \int_B \left[\frac{2+r}{2r} \right] \cos \varphi \sin^2 \varphi dP(r, \varphi) \\ - \int_B \left[\frac{2+r}{2r} \right] \cos \varphi \sin^2 \varphi dP(r, \varphi) & \frac{1}{4} - \int_B \sin^2 \varphi dP(r, \varphi) + I \end{pmatrix} \quad (14)$$

$$\text{mit } I := \int_B \cos^2 \varphi \sin^2 \varphi dP(r, \varphi)$$

zu zeigen, die aus (13) durch Multiplikation der ersten Zeile mit (-1), der zweiten Zeile mit $(-\frac{1}{4})$ und Symmetrisieren entsteht.

Es ist $\int_B \sin^2 \varphi dP(r, \varphi) \leq \frac{1}{4}$ sowie (wegen $\int (x_1 - m(P_1))^2 dP_1(x) \leq 1$)

$q > \frac{1}{2}$ und daher

$$q > \frac{1}{4} + \int_B \sin^2 \varphi dP(r, \varphi) \geq \int_A \frac{1}{4} dP(r, \varphi) + \int_B \sin^2 \varphi dP(r, \varphi).$$

Ferner ist

$$\frac{1}{4} - \int_B \sin^2 \varphi dP(r, \varphi) = \frac{1}{4} \int_A r^2 \sin^2 \varphi dP(r, \varphi).$$

Durch Übergang zu den Integranden ergibt sich dann

- auf A:

$$\begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} r^2 \sin^2 \varphi \end{pmatrix}$$

- auf B:

$$\begin{pmatrix} \sin^2 \varphi + \frac{2}{r} \sin^2 \varphi & - \left[\frac{2+r}{2r} \right] \cos \varphi \sin^2 \varphi \\ - \left[\frac{2+r}{2r} \right] \cos \varphi \sin^2 \varphi & \sin^2 \varphi \cos^2 \varphi \end{pmatrix}$$

Da beide Matrizen P-f.s. positiv definit sind, folgt die Behauptung.

3.3. Teilaussagen im mehrdimensionalen Fall

3.3.1. Bemerkung: Schon in 2.2. und 2.4. hatten wir illustriert, daß die Regularität der kritischen Matrix in den Fixpunkten eine relativ schwache Bedingung ist, die oft an der Grenze zur "Notwendigkeit" liegt. Dies bestätigt sich auch im mehrdimensionalen Fall. Das einzige hierfür bekannte nichttriviale Resultat stammt von Schönholzer (1979) und besagt, daß die Eindeutigkeit auch für Verteilungen in einer Prohorov-Umgebung einer elliptischen Modellverteilung - und daher für die empirische Verteilung hinreichend großer Stichproben - noch gültig bleibt (vgl. auch Hampel e.a. 1986, Remark 3, S. 288). (Dabei sind noch starke Regularitätsvoraussetzungen erforderlich, um u.a. die Differenzierbarkeit auch im Stichprobenfall sicherzustellen.) Der Beweis weist jedoch eine Lücke auf, da es einen Mittelwertsatz der Art, wie er in Schönholzer (1979, S. 38 unten) benutzt wird, im mehrdimensionalen nicht gibt. Es gilt nur das etwa von Ostrowski (1973, Theorem 41.2, S. 266f) angegebene "uniqueness theorem". Die Lücke kann jedoch geschlossen werden, indem man auf den Homotopieansatz 1.3.2.5. zurückgreift und (in der Notation von Schönholzer 1979) für $Q \in N(P, c)$ bzw. die Maße $Q_t := t \cdot Q + (1-t) \cdot P$ die Regularität der kritischen Matrix in den Fixpunkten mittels (7.11) und (7.12) aus Schönholzer (1979) nachweist. Der Homotopieansatz zeigt sich hier erneut als der von den Voraussetzungen her schwächste und bislang einzig erfolgreiche Weg.

Neben Schönholzers Teilresultat für das "volle Problem" sei wieder ein vollständiges Resultat für ein Teilproblem hergeleitet. Wir setzen dazu wieder 3.1.8.(i) voraus.

3.3.2. Situation: Es liege ein lineares Modell der Form $\mathcal{L}(\langle \mu \rangle, \langle \Sigma \rangle)$ vor, d.h. der Erwartungswertunterraum und der Kovarianzmatrizenkegel werden jeweils durch ein (bekanntes) Element $\mu \in \mathbb{R}^p$ und $\Sigma \in PD(p)$ (d.i. positiv definite Matrizen der Ordnung p) aufgespannt. O.B.d.A. kann dabei $\mu = \mu_0 \cdot (1, 0, \dots, 0)^T$ und $\Sigma = \sigma_0^2 \cdot I_p$

angenommen werden. Die Standardschätzer

$$m(P) = \int x_1 dP(x) \quad (15)$$

$$v(P) = \frac{1}{p} \cdot \left[\int (x_1 - m(P))^2 dP(x) + \sum_{i=2}^p \int x_i^2 dP(x) \right]$$

werden in der üblichen Weise robustifiziert, wobei hier die Ellipsoide mit der allgemeinen Konstanten k gebildet werden. (Für $k = \sqrt{p+2}$ ergibt sich wieder das Konzentrationsellipsoid.)

3.3.3. Satz: In der Situation von 3.3.2. gibt es genau einen Fixpunkt (m', v') von d aus (3).

Beweis: Wie im Beweis von 3.2.7. kann man das Problem auf den Nachweis der Regularität der kritischen Matrix $H_{\gamma}(\gamma_0, t_0)$ mit $\gamma_0 = (m_0, v_0) = (0, 1)$ reduzieren. Das Gleichungssystem $H((0, 1), t_0) = 0$ besagt

$$\int_A x_1 dP(x) + \int_B \frac{kx_1}{\sqrt{\sum_{i=1}^p x_i^2}} dP(x) = 0 \quad (16)$$

$$\frac{1}{p} \cdot \left[\sum_{i=1}^p \int_A x_i^2 dP(x) + \int_B k^2 dP(x) \right] = 1$$

mit $A = \{x \in \mathbb{R}^p \mid \|x\|_2 \leq k\}$ und $B = \mathbb{R}^p \setminus A$. Die kritische Matrix ist mit $q := P(A)$:

$$H_{\gamma}(\gamma_0, t_0) = \begin{pmatrix} -q - \int_B \frac{k \cdot \sum_{i=2}^p x_i^2}{\sqrt{\sum_{i=1}^p x_i^2}^3} dP(x) & \int_B \frac{1}{2} \cdot \frac{kx_1}{\sqrt{\sum_{i=1}^p x_i^2}} dP(x) \\ -\frac{1}{p} \cdot \int_A 2x_1 dP(x) & -q - \int_B \frac{p-k^2}{p} dP(x) \end{pmatrix} \quad (17)$$

Nach elementaren Umformungen (Multiplikation mit (-1) , Multiplikation der zweiten Spalte mit 2 und der zweiten Zeile mit $\frac{p}{2}$) und Ausnutzung von (16) gelangt man zu

$$\left[\begin{array}{cc} q + \int_B \frac{k \cdot \sum_{i=2}^p x_i^2}{\sqrt{\sum_{i=1}^p x_i^2}} dP(x) & \int_A x_1 dP(x) \\ \int_A x_1 dP(x) & \sum_{i=1}^p \int_A x_i^2 dP(x) \end{array} \right] \quad (18)$$

Für $p > 1$ ist die Matrix der Integranden von (18) stets p.s.d. und auf einer Menge mit positivem P-Maß sogar p.d. Hieraus folgt die Behauptung. (Der Fall $p = 1$ ist klar.)

3.3.4. Bemerkung (Momentenproblem): Im eindimensionalen Fall konnten Lösungskriterien für das entsprechende Momentenproblem leicht zum Nachweis der Regularität der kritischen Matrix benutzt werden. Über das multidimensionale Momentenproblem ist dagegen nur sehr wenig bekannt (vgl. Fuglede 1983). Es können zwar durchaus hinreichende und notwendige Lösungskriterien angegeben werden (vgl. z.B. McGregor 1980), doch lassen sich diese zumindest zur Lösung des "vollen Problems" nicht leicht zur Anwendung bringen. Dabei reichen im Prinzip Kriterien für die Einheitssphäre und ihren Rand im $\mathbb{R}^p$ aus, da sich das Problem auf zwei "verbundene" Momentenprobleme über der Sphäre bzw. ihrem Rand für Maße mit den Massen q und $1-q$ reduzieren läßt (vgl. die Mengen A und B etwa in 3.2.1.). Im zweidimensionalen Fall reduziert sich das Problem für den Rand der Sphäre sogar auf das "trigonometrische Momentenproblem" (s. z.B. Shohat/Tamarkin 1943, S. 6f); Bedingungen für die Einheitskreisscheibe gibt Atzmon (1975). Diese Untersuchungen behandeln stets das vollständige Momentenproblem (d.h. Momentenfolgen), während mit Blick auf den Nachweis der Regularität von (9) möglichst handliche Kriterien für die ersten zwei (oder vier) Momente benötigt werden. Solche Untersuchungen wären auch aus statistischer Sicht von Interesse, da ja - insbesondere für lineare Modelle - eine zweifache Betrachtungsweise unter Normalverteilungsannahmen und unter Momentenannahmen durchgeführt wird (Witting 1985, S. 452ff; eine besondere Rolle spielt die sogenannte Quasinormalität, vgl. auch Pukelsheim 1977 und Drygas 1972). Der hier eingeschlagene Weg zur Robustifizierung läuft bei Anwendung auf die Standardschätzer auf ein

generalisiertes Momentenproblem hinaus (vgl. 2.2., 2.3. und 2.4.). Er entspricht insofern einer robustifizierten Version der zweiten Betrachtungsweise. Daher war in 3.2.7. und 3.3.3. auch zunächst keine Verteilungsannahme erforderlich. Eine (parametrische) Verteilungsannahme wird erst zur Konsistenzkorrektur benötigt, falls aus der robusten Scattermatrix die ursprüngliche geschätzt werden soll (vgl. auch 2.5.1.(iii) und 2.5.4. (ii)).

Eine andere Möglichkeit, Regularität der kritischen Matrix in einem Fixpunkt zu erhalten, liegt darin, Voraussetzungen aus der Bifurkationstheorie zu machen, die außer der trivialen keine weitere Lösung gestatten (vgl. z.B. Chow/Hale 1982, Chapter 6, S. 215ff). Hierfür ergeben sich aber keine einleuchtenden statistischen Interpretationen.

3.4. Vollständige Aussagen unter abgeschwächten Bedingungen

Über die in 3.2. und 3.3. gewonnenen Aussagen hinaus sollen hier zwei Wege skizziert werden, auf denen Eindeutigkeit auch für das "volle Problem" erreichbar ist.

3.4.1. Pfadverbundene Fixpunkte:

3.4.1.1. Definition: In der Situation von (5) heißt ein Fixpunkt $\gamma(1) = (m', V')$ von W genau dann pfadverbunden zu dem eindeutig bestimmten Fixpunkt $\gamma(0) = (m_0, V_0)$ von Q , wenn eine stetige Abbildung $\gamma: [0,1] \in t \mapsto \gamma(t) \in \Gamma$ existiert derart, daß für jedes $t \in [0,1]$ $H(\gamma(t), t) = 0$ und die Matrix $\pi_\gamma(\partial H(\gamma(t), t))$ regulär ist. γ heißt Verbindungspfad von (m_0, V_0) und (m', V') .

3.4.1.2. Satz: Es gibt höchstens einen Fixpunkt von W , der zu (m_0, V_0) pfadverbunden ist.

Beweis: Falls es zwei Fixpunkte (m_1, V_1) und (m_2, V_2) mit den zugehörigen Verbindungspfaden γ_1 und γ_2 gäbe, so folgte wegen $\gamma_1(0) = \gamma_2(0)$ und $\gamma_1(1) \neq \gamma_2(1)$ für $t' := \inf\{t \in [0,1] \mid \gamma_1(t) \neq \gamma_2(t)\}$ < 1 die Beziehung

$$\det(\pi_\gamma \partial H(\gamma_1(t'), t')) = \det(\pi_\gamma \partial H(\gamma_2(t'), t')) = 0$$

mit Widerspruch zur Voraussetzung.

3.4.1.3. Bemerkung: Pfadverbundenheit lässt sich i.a. "ex post" nachrechnen. Als Ausgangsverteilung wird dabei zweckmäßigerweise diejenige (standardisierte) Verteilung gewählt, die man - ohne Störung - als Modellverteilung ansehen würde. Der Fall empirischer Verteilungen ist direkt mit abgedeckt, da die Vereinbarungen 3.1.8. hier nicht vorausgesetzt wurden.

3.4.1.4. Beispiel: Der Fixpunkt $(0, I_2)$ eines "Orbitalfalles" 3.2.4. W mit der Dichte $h(r, \varphi) = f(r) \cdot g(\varphi)$ ist stets pfadverbunden zum Fixpunkt $(0, I_2)$ der Verteilung Q mit der Dichte $f(r) \cdot \frac{1}{2\pi}$. Die Dichte von $P_t = t \cdot W + (1-t) \cdot Q$ ist nämlich $f(r) \cdot \left[t \cdot g(\varphi) + (1-t) \cdot \frac{1}{2\pi} \right]$. Mit $\gamma(t) \equiv (0, I_2)$ folgt daher die Aussage. Das Beispiel lässt sich auf andere Mittelwerte und Ellipsen verallgemeinern und lässt einen "Orbitalfall" als gestörte elliptische Verteilung erscheinen.

3.4.1.5. Notiz: Pfadverbundenheit lässt sich auch definieren, falls der Fixpunkt von Q in 3.4.1.1. nicht notwendigerweise eindeutig ist. Dann ist jedoch eine entsprechende Voraussetzung in Satz 3.4.1.2. nötig.

3.4.2. Wahl anderer Ellipsen und Projektionen:

Wir behandeln zunächst die "komponentenweise" Anwendung der eindimensionalen Lokations- und Skalenschätzung.

3.4.2.1. Situation: Es liege ein mehrdimensionales Lokations- und Skalenschätzungsproblem vor, wobei die wahre Scattermatrix Σ eine Diagonalmatrix Δ sei. Dann kann man als Ausgangsschätzer komponentenweise die Mittelwerte und Varianz betrachten. Als Norm werde die Maximumsnorm $\|\cdot\|_\infty$, als Ellipsoide die Mengen $\left\{ x \in \mathbb{R}^p \mid \|\sqrt{\Delta}^{-1}(x-m)\|_\infty^2 \leq k^2 \right\}$ (Lokationsfunktion = Identität, Scatterfunktion = symmetrische positiv definite Wurzel, "tuning constant" = k^2) und als Projektion die Projektion in der euklidischen Norm $p_\bullet$ aus 1.2.2.(6) benutzt. Für das Konzentrationsellipsoid ist $k = \sqrt{3}$ zu wählen. Es ist $\gamma = (m, \Delta) \in \mathbb{R}^p \times \{\Delta \mid \Delta \text{ ist p.d. (p \times p)-Diagonalmatrix}\}$. Für die robustifizierte Version der Schätzer gilt dann:

3.4.2.2. Satz: Es sei P ein W -Maß auf $(\mathbb{R}^p, \mathcal{L}^p)$ derart, daß für jede achsenparallele Hyperebene H gilt $P(H) < 1 - \frac{1}{k^2}$. In der Situation aus 3.4.2.1. gibt es dann genau einen Fixpunkt (m', Δ') von d aus (3).

Beweis: Die Aussage folgt unmittelbar durch komponentenweise Anwendung der Ergebnisse aus Abschnitt 2.4.

Obige Situation entspricht formal dem "reduzierten Problem" aus 3.1.6.; doch ist damit das "volle Problem" noch nicht abgedeckt, da die dort benutzte affine Equivarianzeigenschaft für $\|\cdot\|_\infty$ und p_e nicht notwendigerweise gilt. Eine gesonderte Betrachtung ist erforderlich.

3.4.2.3. Situation: Es liege ein mehrdimensionales Lokations- und Skalenschätzproblem vor. Als Ausgangsschätzer werden die Standard-schätzer (vgl. 3.1.1.), als Norm die Maximumsnorm $\|\cdot\|_\infty$, als Ellipsoid zu (m, V) der kleinste achsenparallele Quader $Q(m, V)$ mit $Q(m, V) \supset E(m, V) = \{x \in \mathbb{R}^p \mid \|\sqrt{V}^{-1}(x-m)\|_2^2 \leq k^2\}$ und als Projektion p_e benutzt. Formal ist also $Q(m, V)$ das kartesische Produkt $\prod_{i=1}^p \text{Pr}_i(E(m, V))$, wobei Pr_i die Projektion auf die i -te Komponente ist.

3.4.2.4. Satz: Es sei P ein W -Maß auf $(\mathbb{R}^p, \mathcal{L}^p)$ derart, daß für jede achsenparallele Hyperebene H gilt $P(H) < 1 - \frac{1}{k^2}$. In der Situation aus 3.4.2.3. gibt es dann höchstens einen Fixpunkt (m', V') von d aus (3).

Beweis: Unter Beachtung von Corollary A.9.2.2 aus Mardia/Kent/Bibby (1979, S. 480) ergibt sich

$$Q(m, V) = \prod_{i=1}^p \{m_i - k\sqrt{v_{11}}, m_i + k\sqrt{v_{11}}\}.$$

Daher folgt die Behauptung aus den Ergebnissen von Abschnitt 2.4.

3.4.2.5. Notiz: Die Wahl von $Q(m, V)$ in 3.4.2.3. ist in zweifacher Hinsicht unbefriedigend: zum einen ergeben sich für einen so robustifizierten Schätzer keine Equivarianzeigenschaften, zum anderen ist (bei hoher Korrelation) der Quader unter Umständen erheblich größer als das "erzeugende Ellipsoid" $E(m, V)$ (vgl. auch die entsprechende Problematik für Konfidenzbereiche). Durch das

achsenparallele Heranrücken der außerhalb liegenden Masse können zudem Verzerrungen eintreten. Es ist deshalb näherliegend, die Quader parallel zu den Hauptachsen von $E(m, V)$ zu wählen (im Falle von gleichen Eigenwerten - sogenannten Isotropien - achsenparallel in dem jeweiligen Unterraum). Eine solche Schätzung wäre (falls keine Isotropien auftreten) orthogonal equivariant. Zum Nachweis der Eindeutigkeit kann man aber nicht auf Abschnitt 2.4. zurückgreifen; wir gehen deshalb an dieser Stelle nicht näher darauf ein.

3.4.2.6. Notiz (Zusammenhang mit Influenzkurvenansatz): Die obigen Überlegungen bezogen sich stets darauf, im Stichprobenraum $(X, \mathcal{B})$ selbst zu "winsorisieren". Im Influenzkurvenansatz (Hampel e.a. 1986) geht es dagegen darum, die "scores" $s(x, \vartheta) = \frac{\partial}{\partial \vartheta} \ln f_{\vartheta}(x)$ größenmäßig zu beschränken. Die Bestimmung eines optimalen M-Schätzers mittels einer ψ -Funktion läuft dabei zumindest im "selbststandardisierten Fall" (Hampel e.a. 1986, Section 4.3, S. 238ff) wiederum auf ein mehrdimensionales Lokations- und Skalenschätzproblem hinaus. Die obigen Fixpunktüberlegungen sind also auch in diesem Ansatz von Interesse. Darüber hinaus treten in diesem Ansatz auch Betrachtungen auf, die im wesentlichen eine andere Wahl der Ellipsen und Projektionen bedeuten: Die Partitionierung des Parameterraumes in einen "main part" und einen "nuisance part" (Hampel e.a. 1986, Section 4.4, S. 252ff) macht teilweise Gebrauch von Ellipsen in der Maximumsnorm und der Projektion $p_{\bullet}$. Bei der Untersuchung der Zulässigkeit tritt schließlich die Projektion $p_{\bullet}$ in Verbindung mit den "euklidischen Ellipsen" $E(m, V)$ aus 3.4.2.3. auf (Hampel e.a. 1986, Section 4.6, S. 260ff).

Und, wie das mit Recht geschieht,
Auf die Kunst folgt der Profit.

(W. Busch)

4. WEITERE ANWENDUNGSMÖGLICHKEITEN

In diesem Kapitel sollen einige weitere Anwendungsmöglichkeiten aufgezeigt und damit illustriert werden, daß der in Kapitel 1 entwickelte Ansatz trotz seiner weitreichenden Symmetrievoraussetzungen für viele praktisch wichtige Fälle hilfreich ist. Die Darstellung ist dabei notwendigerweise skizzenhaft; lediglich der Fall der Varianzkomponentenschätzung (Abschnitt 4.3.) wird etwas detaillierter behandelt.

4.1. Robustifizierung einiger klassischer Modelle

Bereits in 3.2.5. wurde angemerkt, daß vielfach robuste Schätzungen wohl deshalb nicht benutzt werden, weil hierzu ein "Umdenken" aus den bekannten Modellen heraus nötig ist. Das Robustifizieren mit Konzentrationsellipsoiden vermeidet dies weitgehend.

4.1.1. Wiederholungsmodell: Das typische Anwendungsgebiet für Robustifizierung mit Konzentrationsellipsoiden ist die n-fache identische unabhängige Wiederholung eines Grundmodells des Typs

$$y = X\beta + \epsilon \quad (1)$$

mit $E(\epsilon) = 0$ und $E(\epsilon\epsilon^T) = D$. Ziel sind Aussagen über β und D . Im allgemeinen wird angenommen, daß $D = D(\theta) = \sum \theta_i D_i$ mit bekannten (meist p.d.) Matrizen D_i . Man bezeichnet derartige Modelle auch als Modelle mit linearer Kovarianzstruktur. Für eine eingehende Behandlung solcher Modelle und der "klassischen" Schätztheorie sei auf Humak (1984) verwiesen; eine Übersicht über die ML-Schätzung gibt Harville (1977). Zu den verschiedenen Schätzern erhält man i.a. auch verschiedene robuste Versionen; es wird hier also nicht ein neuer "robuster Schätzer" für Modell (1) angegeben.

Um die Tragweite des Modells (1) zu demonstrieren, notieren wir einige Spezialfälle:

4.1.2. Spezialfall (Mehrdimensionales Lokations- und Scattermodell): Wählt man in 4.1.1. speziell $X = I_p$, $\beta = \mu$, $D = \Sigma$, so ergibt sich das in Kapitel 3 betrachtete allgemeine Lokations- und Scatterproblem im mehrdimensionalen.

4.1.3. Spezialfall (Correlation Model): In diesem Modell betrachtet man Realisationen unabhängiger Zufallsvariabler $(Y_1, T_1), \dots, (Y_n, T_n)$ wobei Y_i eine skalare Größe und T_i ein $(1 \times s)$ -Zeilenvektor ist. Dabei wird angenommen, daß $n > s+1$ und

$$Y_i | T_i \sim N_1(\delta + T_i \gamma, \sigma^2), \quad T_i^T \sim N_s(\nu, \Sigma), \quad i=1, \dots, n.$$

Als Parameter fungieren $\delta \in \mathbb{R}$, $\gamma \in \mathbb{R}^s$, $\nu \in \mathbb{R}^s$, $\sigma^2 > 0$, $\Sigma \in \text{PD}(s)$. Für eine genaue Diskussion dieses Modells sei auf Arnold (1981, Chapter 16, S. 276ff) verwiesen. Dort wird auch abgeleitet, daß sich dieses Modell mit den Zufallsvektoren $Z_i := (Y_i, T_i)^T$ als Wiederholungsmodell des Grundmodells

$$Z_i \sim N_{s+1} \left(\begin{bmatrix} \delta + \nu^T \gamma \\ \nu \end{bmatrix}, \begin{bmatrix} \sigma^2 + \gamma^T \Sigma \gamma & \gamma^T \Sigma \\ \Sigma \gamma & \Sigma \end{bmatrix} \right)$$

auffassen läßt.

4.1.4. Spezialfall (Verlaufskurven): Anderson (1970, S. 23f) beschreibt eine Anwendung in der Zeitreihenanalyse. Bezeichnet $x_{t\alpha}$ das Ergebnis einer Messung zum Zeitpunkt t ($t = 1, \dots, T$) an einem Objekt α ($\alpha = 1, \dots, N$), so führt die Voraussetzung einer $N_T(\mu, \Sigma)$ -Verteilung für die einzelnen Verlaufskurven $x_\alpha := (x_{1\alpha}, \dots, x_{T\alpha})^T$ unter plausiblen Annahmen zu strukturierten Kovarianzmatrizen Σ . Unterstellt man etwa eine stochastische Differenzgleichung erster Ordnung

$$x_t - \mu_t = \rho(x_{t-1} - \mu_{t-1}) + u_t, \quad t = 2, \dots, T,$$

wobei $u_2, \dots, u_T$ unabhängige $N(0, (1-\rho^2)\sigma^2)$ -verteilte Zufallsvariablen seien und $x_1 - \mu_1$ $N(0, \sigma^2)$ -verteilt sei, so ergibt sich

$$\Sigma = \sigma^2 \cdot \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{bmatrix}.$$

4.1.5. Spezialfall (Circular Stationary Model): Beobachtet man i.i.d. Realisationen eines Zufallsvektors $X = (X_1, \dots, X_p)^T$ mit der Kovarianzstruktur

$$\text{Var}(X_i) = \sigma^2$$

$$\text{Cor}(X_i, X_{i+k}) = \sigma^2 \rho_k \quad \text{mit } \rho_k = \rho_{p-k}$$

für $i = 1, \dots, p$, $k = 1, \dots, p-1$ und $2 \leq i+k \leq p$, so spricht man von einem zirkulär stationären Modell (Olkin/Press 1969). Derartige Modelle treten bei der Untersuchung von Symmetrien (z.B. in der Biometrie) auf (vgl. z.B. Olkin 1974).

4.2. Indirekte Robustifizierung

In den unter 4.1. genannten Beispielen war das Konzentrationsellipsoid stets "direkt" aus dem Modell heraus erkennbar. In anderen Fällen schätzt man das robuste Konzentrationsellipsoid als Zwischenstufe, um daraus abgeleitete Größen zu robustifizieren.

4.2.1. Korrelationskoeffizientenschätzung: Während man versuchen kann, die Kovarianzmatrix aus robusten Schätzungen für die Korrelationen aufzubauen (Huber 1981, Section 8.3, S. 204ff), werden bei der Robustifizierung des Korrelationskoeffizienten Lokation und Scatter nur "nebenher" geschätzt. Zur robusten Schätzung des Korrelationskoeffizienten ρ zweier Zufallsvariablen X und Y mit der gemeinsamen Verteilung P , berechnet man also zunächst die robustifizierten Lokations- und Scatterwerte

$$m(P) = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} \text{ und } V(P) = \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix} \quad (2)$$

und hieraus als robusten Schätzwert für ρ den Wert $\hat{\rho} = \frac{v_{12}}{\sqrt{v_{11}}\sqrt{v_{22}}}$. Vorinformation (z.B. die Kenntnis von $\frac{v_{11}}{v_{22}}$ o.ä.) wird man bei der Schätzung von (2) berücksichtigen. Da $V(P)$ nicht zur Schätzung der Kovarianzmatrix von P benutzt werden soll, ist eine Konsistenzkorrektur - und damit eine Verteilungsannahme - nicht erforderlich. Gegenüber reiner Ausreißerbehandlung (z.B. koordinatenweisem trimming) scheint die Berücksichtigung der multivariaten Struktur bei der Korrelationskoeffizientenschätzung vorteilhaft (vgl. Bebbington 1978, S. 221). Das obige Verfahren läßt sich als ein "ellipsoidal Winsorizing" ansehen; wie das "ellipsoidal trimming" dürfte es vor allem für kontaminierte elliptische Verteilungen relativ gute Ergebnisse zeigen (vgl. Titterton 1978, S. 227). Mit welcher Art von Kontamination man zu rechnen hat, wird allerdings häufig nur in speziellen Modellen (z.B. für Innerklassenkorrelationskoeffizienten) genauer festlegbar sein; für einen solchen Fall werden in Abschnitt 4.3. einige Simulationsergebnisse angegeben.

4.2.2. Einfache lineare Regression: Diese Modelle werden in der statistischen Praxis sehr häufig benutzt; hier wird der Fall der Regression mit stochastischen Regressoren betrachtet. Für die dabei vorliegenden reellen Zufallsvariablen X und Y und dem durch

$$m = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} \text{ und } \Sigma = \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix} \quad (3)$$

gegebenen Konzentrationsellipsoid bestehen dann folgende Zusammenhänge (Cramér 1946, S. 284f):

- (i) Die Gerade durch m und $(m_1 + k \cdot \sigma_1, m_2 + \frac{\rho \cdot k}{\sigma_1})^T$ ist die Gerade der Regression von Y auf X .
- (ii) Die Gerade durch m und $(m_1 + \frac{\rho \cdot k}{\sigma_2}, m_2 + k \cdot \sigma_2)^T$ ist die Gerade der Regression von X auf Y .
- (iii) $\lambda_1 \geq \lambda_2 > 0$ seien die Eigenwerte von Σ . Im Fall $\lambda_1 > \lambda_2$ ist die durch die erste Hauptachse (Eigenvektor zu λ_1 durch m)

festgelegte Gerade die Gerade der orthogonalen Regression von X und Y . Im Fall $\lambda_1 = \lambda_2$ ist die orthogonale Regression nicht bestimmt.

- (iv) Durch Veränderung der Konstanten k^2 (für die Bildung des Ellipsoids) erhält man eine Schar homothetischer Ellipsen mit dem gemeinsamen Mittelpunkt m .

Die robuste Schätzung von (3) liefert also direkt robuste Versionen für die drei genannten Regressionen. Dazu sind noch die folgenden Anmerkungen zu machen:

1. Wegen (iv) fungiert k^2 wieder als "tuning constant", mit der man den Grad der Robustifizierung wählt.
2. Es besteht die bekannte Identität mit den Trägheitsellipsen (Cramér 1946, S. 275f).
3. Vorhandenes Wissen über die Struktur von (3) wird man bei der Robustifizierung benutzen. Wiederum kann man durch verschiedene Ellipsen verschiedene Effekte erzielen. Dies entspricht der Vielfalt der Ansätze zur robusten Regression (vgl. auch die sehr anschauliche Darstellung in Hampel e.a. 1986, Figure 1, S. 332).
4. Falls X nicht stochastisch ist, besteht dennoch Veranlassung, den Einfluß großer x -Werte zu vermindern, um sogenannte "leverage points" (Huber 1981, S. 155) zu vermeiden.
5. Regression wird häufig benutzt, um "eine Gerade an eine Punktwolke anzupassen". Hierfür scheint - schon aus Symmetriegründen - (iii) besonders geeignet zu sein. In der Tat übernehmen letztere Modelle (unter der Bezeichnung "linear statistical relationship models") in neuerer Zeit mehr und mehr die Rolle der hierfür oft unzutreffenderweise benutzten "klassischen" Regression. Der obige Ansatz liefert also zugleich einen Vorschlag zur Robustifizierung solcher Modelle (vgl. Anderson 1984, S. 6ff).
6. Der Umstand, daß hier Regression letztlich via einer Lokations- und Scatterschätzung robustifiziert wird, muß nicht gegen den Ansatz sprechen (vgl. die Bemerkung in Huber 1981, S. 195).
7. Die Aussagen lassen sich auf multiple Regression verallgemeinern (vgl. Cramér 1946, S. 309f).

4.3. Robustifizierung von Varianzkomponentenmodellen

4.3.1. Definition:

- (a) Ein lineares Modell $\mathcal{L}(\mathcal{M}, \mathcal{V})$ liegt vor, falls für einen $\mathbb{R}^p$ -wertigen Zufallsvektor Y die Beziehungen $E(Y) \in \mathcal{M}$ und $\text{Cov}(Y) \in \mathcal{V}$ gelten, wobei $\mathcal{M}$ ein linearer Unterraum von $\mathbb{R}^p$ und $\mathcal{V}$ ein Kegel (symmetrischer) p.s.d. (p x p)-Matrizen ist.
- (b) Falls $\mathcal{M}$ durch eine Basis $\{m_1, \dots, m_1\}$ in der Form

$$\mathcal{M} = \left\{ \sum_{i=1}^1 \mu_i m_i \mid \mu_i \in \mathbb{R} \right\}$$

und $\mathcal{V}$ durch ein Erzeugendensystem $\{V_1, \dots, V_k\}$ p.s.d. Matrizen in der Form

$$\mathcal{V} = \left\{ \sum_{i=1}^k \sigma_i^2 \cdot V_i \right\}$$

vorliegt, heißt das lineare Modell Varianzkomponentenmodell. Die Koeffizienten σ_i^2 (≥ 0) heißen auch Varianzkomponenten.

4.3.2. Bemerkungen:

- (i) Ein Modell mit gemischten Effekten ist ein Modell der Form

$$Y = X\beta + \sum_{i=1}^k U_i u_i, \quad (4)$$

wobei

Y ein (beobachtbarer) p-dimensionaler Zufallsvektor

X eine (bekannte) (p x ℓ)-Matrix

β ein (unbekannter) Parametervektor aus $\mathbb{R}^\ell$

U_i (bekannte) (p x p_i)-Matrizen

u_i (unbeobachtbare) p_i -dimensionale Zufallsvektoren mit

$E(u_i) = 0$, $\text{Cov}(u_i) = \sigma_i^2 \cdot D_i$ mit bekanntem p.d. D_i und

$E(u_i u_j^T) = 0 \quad \forall i, j$

sind. Dies ist gerade ein lineares Modell, bei dem $\mathcal{V}$ durch die Matrizen $U_i D_i U_i^T =: V_i$ gegeben ist. Ist speziell

$D_i = I_{p_i}$, so ergibt sich ein Varianzkomponentenmodell.

- (ii) Formel (4) deutet an, in welchen Zusammenhängen Varianzkomponentenmodelle auftreten. Die Matrizen X und U_i sind häufig Designmatrizen (d.h. nur aus den Elementen 0 und 1 aufgebaut).
- (iii) Die Bezeichnungen sind in der Literatur nicht einheitlich;

gelegentlich werden auch Modelle als Varianzkomponentenmodelle zugelassen, bei denen zwar $\sum \theta_1 V_1$ p.s.d., aber nicht notwendigerweise $\theta_1 \geq 0$ und V_1 p.s.d. sind.

4.3.3. Spezialfall (one-way random effects model): Der balanzierte Fall der Ein-Weg-Klassifikation ist der einfachste Spezialfall eines Varianzkomponentenmodells. Er ergibt sich durch die Modellgleichung

$$Y_{ij} = \mu + a_i + e_{ij} \quad , \quad i=1, \dots, n; \quad j=1, \dots, p. \quad (5)$$

Dabei sind a_i und e_{ij} unabhängige (unbeobachtbare) Zufallsvariablen mit den Verteilungen $a_i \sim N(0, \sigma_a^2)$ und $e_{ij} \sim N(0, \sigma_e^2)$. Als Parameter fungieren $-\infty < \mu < \infty$, $\sigma_a^2 \geq 0$ und $\sigma_e^2 > 0$. Die Kovarianzstruktur ist gegeben durch

$$\text{Cov}(Y_{ij}, Y_{kl}) = \begin{cases} \sigma_a^2 + \sigma_e^2, & i = k, \quad j = l \\ \sigma_a^2, & i = k, \quad j \neq l \\ 0, & i \neq k \end{cases} \quad (6)$$

Mit der Definition $Y_i := (Y_{i1}, \dots, Y_{ip})^T$ besteht (5) also aus (Realisationen von) n i.i.d. Zufallsvektoren Y_i mit

$$Y \sim N_p \left[\begin{bmatrix} \mu \\ \cdot \\ \cdot \\ \cdot \\ \mu \end{bmatrix}, \begin{bmatrix} \sigma_a^2 + \sigma_e^2 & \sigma_a^2 & \dots & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 + \sigma_e^2 & & \sigma_a^2 \\ & & \ddots & \\ \sigma_a^2 & & & \sigma_a^2 + \sigma_e^2 \end{bmatrix} \right] \quad (7)$$

4.3.4. Bemerkung (Modelldiskussion): Für eine genauere Diskussion des Modells 4.3.3. sei auf Arnold (1981, Chapter 15, S. 242ff) verwiesen. Wir halten hier fest, daß (7) gerade die zugrundeliegende multivariate Struktur beschreibt. Die Kovarianzmatrix läßt sich in der Form

$$\Sigma = (\sigma_a^2 + \sigma_e^2) \cdot [(1-\rho)I_p + \rho J_p] \quad (8)$$

schreiben, wobei $J_p = \frac{1}{p} \cdot 1_p^T$ eine Matrix, deren sämtliche Elemente

1 sind, und ρ den Innerklassenkorrelationskoeffizienten $\frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$

bezeichnet. Ohne den Vorfaktor $(\sigma_a^2 + \sigma_e^2)$ bezeichnet man eine Matrix des Typs (8) auch als Equikorrelationsmatrix. Die Matrix (8) hat die Eigenwerte $\lambda_1 = (\sigma_a^2 + \sigma_e^2) \cdot (1 + (p-1) \cdot \rho)$ und $\lambda_2 = \dots = \lambda_p = (\sigma_a^2 + \sigma_e^2) \cdot (1 - \rho)$ und als zugehörige Eigenvektoren 1_p und $(p-1)$ zu 1_p orthogonale Eigenvektoren (Mardia/Kent/Bibby 1979, S. 471f). Das durch Σ aus (8) vermittelt

$$(x - \mu \cdot 1_p)^T \Sigma^{-1} (x - \mu \cdot 1_p) \leq k^2 \quad (9)$$

festgelegte Ellipsoid besitzt daher den Mittelpunkt $\mu \cdot 1_p$, die Hauptachse 1_p mit der Länge $k \cdot \sqrt{p\sigma_a^2 + \sigma_e^2}$ und $(p-1)$ zu 1_p orthogonale isotrope Nebenachsen mit den Längen $k \cdot \sqrt{\sigma_e^2}$ (Mardia/Kent/Bibby 1979, S. 484f). Die Matrix (8) ist wegen $0 \leq \rho < 1$ stets p.d. Das durch (9) gegebene Ellipsoid wird im Fall $\sigma_a^2 = 0$ zu einer Kugel. In diesem Fall sind alle a_i in (5) gleich 0.

4.3.5. Bemerkung (Schätzer für $\mu, \sigma_e^2, \sigma_a^2$): Übersichtlicher Ausgangspunkt für die statistische Behandlung des Modells 4.3.3. ist die sogenannte ANOVA-Tafel:

Source	DF	SS	MS	E(MS)
between	n-1	SSA	$MSA := \frac{SSA}{n-1}$	$p \cdot \sigma_a^2 + \sigma_e^2$
within	n(p-1)	SSE	$MSE := \frac{SSE}{n(p-1)}$	σ_e^2
total	np-1	SST = SSA + SSE		

Dabei ist $SSA = p \cdot \sum_{i=1}^n (\bar{Y}_{i.} - \bar{Y}_{..})^2$ und $SSE := \sum_{i=1}^n \sum_{j=1}^p (Y_{ij} - \bar{Y}_{i.})^2$. $(\bar{Y}_{..}, SSA, SSE)$ ist eine vollständige suffiziente Statistik für Modell 4.3.3. Wir notieren drei wesentliche Schätzer für $(\mu, \sigma_e^2, \sigma_a^2)$ und verweisen bezüglich der Herleitung auf Arnold (1981, S. 246ff) und Hartung (1981):

- ANOVA-Schätzer:

$$\hat{\mu} = \bar{Y}_{..}$$

$$\hat{\sigma}_e^2 = MSE$$

$$\hat{\sigma}_a^2 = \frac{1}{p} \cdot (MSA - MSE)$$

Dies sind die UMVU-Schätzer im Modell 4.3.3. $\hat{\sigma}_a^2$ ist nicht notwendigerweise nichtnegativ; es existiert jedoch kein nichtnegativer unverfälschter Schätzer.

- ML-Schätzer:

$$\hat{\mu} = \bar{Y}_{..}$$

$$\hat{\sigma}_e^2 = \text{Min} \left\{ MSE, \frac{1}{np} \cdot (SSA + SSE) \right\}$$

$$\hat{\sigma}_a^2 = \text{Max} \left\{ 0, \frac{1}{np} \cdot (SSA - \frac{1}{p-1} \cdot SSE) \right\}$$

- Hartung-Schätzer:

$$\hat{\mu} = \bar{Y}_{..}$$

$$\hat{\sigma}_e^2 = MSE$$

$$\hat{\sigma}_a^2 = \frac{p}{1+p^2} \cdot MSA$$

Dieser Schätzer minimiert den bias unter der Bedingung der Nichtnegativität.

Einen Vergleich verschiedener Schätzer mit Bezug auf den quadratischen Verlust findet man bei Klotz/Milton/Zacks (1969).

4.3.6. Bemerkung (Transformation): Die orthogonale Matrix H_p besitze die Zeilen

$$H_p^{(1)} = \left[\frac{1}{\sqrt{p}}, \dots, \frac{1}{\sqrt{p}} \right]$$

$$H_p^{(j)} = (h_j, \dots, h_j, -(j-1)h_j, 0, \dots, 0) \text{ für } j = 2, \dots, p$$

$$\left[\text{wobei } h_j = \frac{-1}{\sqrt{j(j-1)}} \text{ j-mal wiederholt wird} \right],$$

d.h. H_p ist die Helmert-Matrix der Ordnung p, die mit Ausnahme der ersten Zeile mit (-1) multipliziert wurde. Durch Anwendung der Transformation

$$Z_1 = H_p \cdot Y_1$$

erhält man aus (7) Zufallsvektoren Z_1 mit

$$E(Z_1) = \begin{bmatrix} \sqrt{p} \mu \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{und} \quad \text{Cov}(Z_1) = \begin{bmatrix} p\sigma_a^2 + \sigma_e^2 & 0 & \dots & 0 \\ 0 & \sigma_e^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sigma_e^2 \end{bmatrix}. \quad (10)$$

Zur Robustifizierung verzichten wir zunächst wieder auf die in 4.3.3. gemachte Normalverteilungsannahme und gehen direkt von (10) aus. Wir robustifizieren hier den ANOVA-Schätzer und drücken diesen daher zunächst in den Z_i aus. Man erhält

$$\begin{aligned} \bar{y}_{..} &= \frac{1}{\sqrt{p}} \cdot \bar{z}_{11} \\ \text{SSA} &= \sum_{i=1}^n (Z_{11} - \bar{z}_{11})^2 \\ \text{SSE} &= \sum_{i=1}^n \sum_{j=2}^p Z_{ij}^2 \end{aligned} \quad (11)$$

Das Modell (10) läßt sich bequemer in den Parametern $m_0 := \sqrt{p} \mu$, $v_{11} := p\sigma_a^2 + \sigma_e^2$ und $v_{22} := \sigma_e^2$ schreiben als

$$E(Z_1) = \begin{bmatrix} m_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} =: m \quad \text{und} \quad \text{Cov}(Z_1) = \begin{bmatrix} v_{11} & 0 & 0 \\ 0 & v_{22} & \vdots \\ \vdots & & \ddots \\ 0 & 0 & v_{22} \end{bmatrix} =: V. \quad (12)$$

Die ANOVA-Schätzer sind dann

$$\begin{aligned} \hat{m}_0 &= \bar{z}_{11} = \frac{1}{n} \sum_{i=1}^n Z_{11} \\ \hat{v}_{11} &= \text{MSA} = \frac{1}{n-1} \sum_{i=1}^n (Z_{11} - \bar{z}_{11})^2 \\ \hat{v}_{22} &= \text{MSE} = \frac{1}{p-1} \sum_{j=2}^p \left[\frac{1}{n} \sum_{i=1}^n Z_{ij}^2 \right]. \end{aligned} \quad (13)$$

Mit der Notation $z_{(i)}$ für die i -te Komponente eines Vektors z läßt sich dies (bei Ersetzen des Nenners $n-1$ durch n in der Formel für $\hat{v}_{11}$) in folgender Weise als Funktional schreiben

$$m_0(P) = \int z_{(1)} dP(z)$$

$$v_{11}(P) = \int (z_{(1)} - m_0(P))^2 dP(z) \quad (14)$$

$$v_{22}(P) = \frac{1}{p-1} \cdot \int \sum_{j=2}^p z_{(j)}^2 dP(z).$$

Entsprechende funktionale Analoga erhält man mittels (13) und (14) auch für den ML-Schätzer und den Hartung-Schätzer. Die Formeln (14) zeigen, daß der ANOVA-Schätzer im Grunde nur einfache Varianzen und Mittelwerte schätzt. Negative Schätzungen für σ_a^2 ergeben sich genau dann, wenn das durch $\text{Cov}(Z)$ gegebene Ellipsoid als "stehend" statt als "liegend" geschätzt wird.

4.3.7. Situation: Es sei W ein W -Maß mit einer λ^p -f.ü. stetigen λ^p -Dichte. Wir wollen (14) mit Hilfe der durch m und V aus (12) gegebenen Ellipsoide robustifizieren und gehen dazu wie in Kapitel 3 vor. Mit $P_t := t \cdot W + (1-t) \cdot Q$ und der Funktion

$$\psi(z) = \begin{cases} z & , \text{ falls } \|V^{-1}(z-m)\|_2^2 \leq p+2 \\ \frac{\sqrt{p+2}}{\|V^{-1}(z-m)\|_2} (z-m) + m & , \text{ falls } \|V^{-1}(z-m)\|_2^2 > p+2 \end{cases} \quad (15)$$

ergibt sich für die Funktion $H(\gamma, t) = H((m_0, v_{11}, v_{22}), t)$:

$$H((m_0, v_{11}, v_{22}), t) = \begin{cases} \int \psi(z)_{(1)} dP_t(z) - m_0 \\ \int [\psi(z)_{(1)} - m_0]^2 dP_t(z) - v_{11} \\ \frac{1}{p-1} \cdot \int \sum_{j=2}^p [\psi(z)_{(j)}]^2 dP_t(z) - v_{22} \end{cases} \quad (16)$$

4.3.8. Satz: Es sei $p = 2$ und $((0, 1, 1), t_0)$ eine Nullstelle von H aus (16). Bei Verwendung von Polarkoordinaten und der Notation P für das transformierte Maß P_{t_0} gilt dann in der Situation von

4.3.7. mit A , B und q wie in 3.2.1.:

- (i) Das Gleichungssystem $H((0, 1, 1), t_0) = 0$ entspricht den Gleichungen (8.a), (8.c) und (8.d) aus 3.2.1.(i).
- (ii) Die kritische Matrix $H_\gamma((0, 1, 1), t_0)$ erhält man aus der Matrix

(9) aus 3.2.1.(ii) durch Streichen der zweiten Spalte und der zweiten Zeile.

Beweis: Die Situation entspricht gerade dem Fall, daß $m_2 = 0$ in 3.2.1. bekannt ist.

4.3.9. Notiz: Obwohl die kritische Matrix aus 4.3.8.(ii) eine Vereinfachung gegenüber (9) aus 3.2.1.(ii) ist und zwischen den Fällen 3.2.7. und 3.2.1. liegt, ist ein allgemeiner Nachweis der Regularität nicht gelungen. Das ändert sich auch nicht, wenn man statt der Funktion H aus (16) die leicht veränderte Funktion $\tilde{H}$ mit der zweiten Komponente

$$\int \left[\psi(z)_{(1)} \right]^2 dP_t(z) - m_0^2 - v_{11}$$

untersucht. $\tilde{H}$ besitzt dieselben Nullstellen wie H ; die kritische Matrix vereinfacht sich aber nicht. Bei Vorliegen genauerer a-priori Schranken für (m_0, v_{11}, v_{22}) ist der Regularitätsnachweis aber u.U. zu erbringen. Einige interessante Teilresultate ergeben sich bereits aus den Untersuchungen von Kapitel 3:

4.3.10. Bemerkung (Eindeutigkeit in Spezialfällen): Es liege Situation 4.3.7. vor.

(i) Es sei $p = 2$ und $v_{11} = 2\sigma_a^2 + \sigma_e^2$ bekannt und o.E. gleich 1.

Dann gibt es in der Situation 3.2.6. nach 3.2.7. genau einen Fixpunkt (m'_0, v'_{22}) .

(ii) Es sei $\frac{\sigma_a^2}{\sigma_e^2} = c$ bekannt und daher $v_{11} = (pc + 1) \cdot v_{22}$. Dann gibt es nach 3.3.3. genau einen Fixpunkt (m'_0, v'_{22}) .

(iii) Transformiert man (10) durch Addition der zweiten bis zur p -ten Komponente auf das Modell

$$\begin{pmatrix} \sqrt{p} \mu \\ 0 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} p\sigma_a^2 + \sigma_e^2 & 0 \\ 0 & (p-1)\sigma_e^2 \end{pmatrix},$$

so erhält man in den Parametern $m_0 = \sqrt{p} \mu$, $v_{11} = p\sigma_a^2 + \sigma_e^2$ und $v_{22} = (p-1)\sigma_e^2$ durch komponentenweise Robustifizierung einen eindeutigen Fixpunkt (m'_0, v'_{11}) in der ersten Komponente und einen eindeutigen Fixpunkt v'_{22} in der zweiten Komponente, falls die entsprechenden Bedingungen aus Abschnitt 2.4.

bzw. Abschnitt 2.3. erfüllt sind. Die Voraussetzungen einer λ^P -f.ü. stetigen Dichte für P ist in diesem Fall nicht erforderlich.

4.3.11. Anwendungen: Trotz seiner einfachen Struktur findet das Ein-Weg-Klassifizierungsmodell vielfache Anwendungen:

(i) Statistische Genetik: Hier ist man insbesondere an dem

Innerklassenkorrelationskoeffizienten $\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$ interessiert,

der auch als Heritabilitätskoeffizient bezeichnet wird. ρ gibt den Anteil der genetischen Varianz an der Gesamtvarianz eines Merkmals an und ist somit von großer Bedeutung für die Beurteilung von Züchtungsmaßnahmen (z.B. Selektion).

(ii) Technik: Das Modell 4.3.3. tritt in vielen Abwandlungen im Ingenieurbereich auf, z.B. in der Geodäsie. Ein Spezialfall aus der Meßtechnik ist etwa die Analyse von Ringversuchen. Hier tritt wieder der Innerklassenkorrelationskoeffizient in einer ähnlichen Bedeutung wie in (i) auf.

4.3.12. Einige Simulationsresultate: Einen Eindruck von der Genauigkeit der Schätzungen kann man sich mittels Monte Carlo Studien verschaffen. Wir müssen uns an dieser Stelle auf einige erste Ergebnisse beschränken, die sich an folgenden Überlegungen orientieren:

1. Zielgröße sind das Gesamtmittel μ und der Innerklassenkorrelationskoeffizient ρ . Dies erfordert keine Verteilungsannahme bzw. Konsistenzkorrektur, deckt aber schon einige wesentliche Fälle ab (vgl. 4.3.11. und auch 4.2.1.). Im Gegensatz zu weitgehend willkürlich festgelegter multivariater Kontamination, wird man hier - der Genese des Modells folgend - Kontamination der Verteilung der zufälligen Effekte a und der Zufallsfehler e annehmen. Als Schätzer werden die ANOVA-Schätzer in der ursprünglichen Form aus 4.3.5. benutzt. Ausgehend von den unrobustifizierten Werten werden durch einfache Iteration die robustifizierten Schätzwerte errechnet. Dies Verfahren ist konzeptionell und programmiertechnisch insofern bequem, als jeweils die klassischen Schätzer berechnet werden, wobei die Daten jeweils durch Winsorisieren am Konzentrationsellipsoid (temporär) modifiziert werden. Das Verfahren wird ab-

gebrochen, sobald

$$|\hat{\mu}^{(1+1)} - \hat{\mu}^{(1)}| + |\hat{\sigma}_a^{2(1+1)} - \hat{\sigma}_a^{2(1)}| + |\hat{\sigma}_e^{2(1+1)} - \hat{\sigma}_e^{2(1)}| < 10^{-9}$$

oder $i = 200$ ist. Im letzteren Fall gilt $(\hat{\mu}^{(201)}, \hat{\sigma}_a^{2(201)}, \hat{\sigma}_e^{2(201)})$ als Lösung.

2. Es werden stets 100 Wiederholungen durchgeführt und Mittelwert und Standardabweichung berechnet.
3. Das Design wird so gewählt, daß i.a. $p \cdot n \approx 100$, um die Resultate vergleichbar zu halten. Neben leicht "unterquadratischen" (d.h. $n < p$) und "quadratischen" (d.h. $n = p$) Designs wie sie in 4.3.11.(i) auftreten, werden insbesondere auch stark "überquadratische" (d.h. $n > p$) Designs betrachtet, weil diese im Fall 4.3.11.(ii) besonders realistisch sind.
4. Grundsätzlich wird versucht, möglichst viele "Quervergleiche" zu ermöglichen; genaueres ergibt sich aus dem Aufbau der Studie.

Die Tabelle 4.1. (s. S. 104) gibt die Ausgangsdaten für die insgesamt 20 Simulationsläufe an. Dabei bedeuten:

VA: Verteilung der zufälligen Effekte a

VE: Verteilung der Zufallsfehler e

Startwert: Wert zur Initialisierung des Zufallszahlengenerators.

Wir verwenden weiter die Notation $N(a,b)$ für die Normalverteilung mit Mittelwert a und Varianz b sowie $C(a,b)$ für die Cauchy-Verteilung mit Median a und Semi-Interquartile Range $\sqrt{b}$. Zur Erzeugung der Zufallszahlen wurden die Programme G05CAF, G05DDF und G05DFF aus der NAG-Library benutzt.

Die Simulationen wurden so angesetzt, daß mit vier Ausnahmen als Zielwerte (d.h. als "wahre" Werte ohne Berücksichtigung der Kontaminationen) $\mu = 0.0$ und $\rho = 0.5$ zugrunde lagen. Die Ausnahmen sind:

C5 mit $\mu = 0.0$ und $\rho = 0.3$

C6 mit $\mu = 0.0$ und $\rho = 0.8$

C9 mit $\mu = 1.0$ und $\rho = \frac{1}{3}$

C10 mit $\mu = 1.0$ und $\rho = \frac{2}{3}$.

Tabelle 4.1.: Ausgangsdaten für die Simulationen

Kürzel	n	p	Verteilungen VA und VE	Startwert
N1	10	10	$VA = N(0,1) = VE$	7391
N2	25	4	$VA = N(0,1) = VE$	985104
N3	12	8	$VA = N(0,1) = VE$	28
N4	8	12	$VA = N(0,1) = VE$	65577
T1	10	10	$VA = 0.9 \cdot N(0,1) + 0.1 \cdot N(0,9) = VE$	23887
T2	25	4	$VA = 0.9 \cdot N(0,1) + 0.1 \cdot N(0,9) = VE$	22771
T3	12	8	$VA = 0.9 \cdot N(0,1) + 0.1 \cdot N(0,9) = VE$	81112
T4	8	12	$VA = 0.9 \cdot N(0,1) + 0.1 \cdot N(0,9) = VE$	376
C1	10	10	$VA = 0.9 \cdot N(0,1) + 0.1 \cdot C(0,1) = VE$	9932
C2	25	4	$VA = 0.9 \cdot N(0,1) + 0.1 \cdot C(0,1) = VE$	12390
C3	12	8	$VA = 0.9 \cdot N(0,1) + 0.1 \cdot C(0,1) = VE$	92887
C4	8	12	$VA = 0.9 \cdot N(0,1) + 0.1 \cdot C(0,1) = VE$	639920
C5	10	10	$VA = 0.9 \cdot N(0,0.3) + 0.1 \cdot C(0,0.3)$ $VE = 0.9 \cdot N(0,0.7) + 0.1 \cdot C(0,0.7)$	28762
C6	10	10	$VA = 0.9 \cdot N(0,2.0) + 0.1 \cdot C(0,2.0)$ $VE = 0.9 \cdot N(0,0.5) + 0.1 \cdot C(0,0.5)$	38827
C7	10	10	$VA = 0.95 \cdot N(0,1) + 0.05 \cdot C(0,1)$ $VE = 0.85 \cdot N(0,1) + 0.15 \cdot C(0,1)$	881900
C8	10	10	$VA = 0.85 \cdot N(0,1) + 0.15 \cdot C(0,1)$ $VE = 0.95 \cdot N(0,1) + 0.05 \cdot C(0,1)$	993
CBA	10	10	$VA = 0.9 \cdot N(0,1) + 0.1 \cdot C(1,1)$ $VE = 0.9 \cdot N(0,1) + 0.1 \cdot C(0,1)$	52677
CBE	10	10	$VA = 0.9 \cdot N(0,1) + 0.1 \cdot C(0,1)$ $VE = 0.9 \cdot N(0,1) + 0.1 \cdot C(1,1)$	115523
C9	20	4	$VA = 0.99 \cdot N(0,1) + 0.01 \cdot C(0,1)$ $VE = 0.95 \cdot N(0,2) + 0.05 \cdot C(0,2)$	27651
C10	25	2	$VA = 0.98 \cdot N(0,2) + 0.02 \cdot C(0,2)$ $VE = 0.95 \cdot N(0,1) + 0.05 \cdot C(0,1)$	198276

Die Tabelle 4.2. (s. S. 106) zeigt die Resultate der Simulationen. Dabei bedeuten:

- MU: Mittelwert (über 100 Replikationen) der Schätzung für μ
RMU: Mittelwert (über 100 Replikationen) der robustifizierten Schätzung für μ
RHO: Mittelwert (über 100 Replikationen) der Schätzung für ρ
RRHO: Mittelwert (über 100 Replikationen) der robustifizierten Schätzung für ρ
ANEG: Anzahl (auf 100, d.h. in %) von negativen Schätzwerten für ρ (bzw. σ_a^2)
RANEG: Anzahl (auf 100, d.h. in %) von negativen robustifizierten Schätzwerten für ρ (bzw. σ_a^2)
MAXIT: Maximale Anzahl von Iterationen.

Die Zahlen in Klammern geben die entsprechenden Standardabweichungen für MU, RMU, RHO und RRHO an. Das Ergebnis zeigt im Rahmen der mit einer Studie dieses Umfangs erzielbaren Aussagekraft folgendes:

- (i) Bei den Normalverteilungsmodellen N1-N4 wird durch Robustifizierung kaum Effizienz verloren. Bei "Tukey-Kontaminationsmodellen" und insbesondere bei "Cauchy-Kontaminationsmodellen" ist Robustifizierung aber sehr vorteilhaft. Dies trifft für alle untersuchten Kombinationen von n und p zu.
- (ii) Die Läufe C5 und C6 deuten an, daß dies nicht nur für den Zielwert $\rho=0.5$, sondern auch für $\rho=0.3$ und $\rho=0.8$ gilt. Für die weiteren Simulationen wurde daher weiterhin der in der Mitte gelegene Wert $\rho=0.5$ benutzt. Dies soll nicht implizieren, daß der Grenzfall $\sigma_a^2 = \sigma_e^2$ am häufigsten auftritt.
- (iii) Mit den Simulationen C7 und C8 wird das Verhalten bei verschieden starken Kontaminationen von VA und VE illustriert. Während Robustifizierung in beiden Fällen sinnvoll ist, scheint sie die Resultate im ersten Fall besonders stark zu verbessern.
- (iv) Die Läufe CBA und CBE illustrieren den Einfluß eines zusätzlichen bias bei VA bzw. VE.
- (v) Die Simulationen C9 und C10 sind Beispiele für "small sample Fälle" (im Sinne von $p \cdot n < 100$) zu extrem überquadratischen Fällen, die in den Anwendungen häufig sind (z.B. Ringversuche). Aber auch in der Biometrie ist oft p klein (z.B. in der "Zwillingsforschung"; vgl. Smith/Murray 1984

Tabelle 4.2.: Ergebnisse der Simulationen

Kürzel	MU	RMU	RHO	RRHO	ANEG	RANEG	MAXIT
N1	-0.04 (0.32)	-0.04 (0.32)	0.47 (0.13)	0.47 (0.13)	0	0	107
N2	0.02 (0.24)	0.02 (0.25)	0.47 (0.12)	0.47 (0.11)	0	0	43
N3	0.02 (0.30)	0.02 (0.30)	0.47 (0.15)	0.47 (0.15)	0	0	63
N4	0.01 (0.37)	0.01 (0.38)	0.47 (0.14)	0.48 (0.14)	0	0	160
T1	-0.04 (0.49)	-0.03 (0.48)	0.47 (0.19)	0.50 (0.18)	0	0	127
T2	0.01 (0.24)	0.01 (0.23)	0.45 (0.14)	0.48 (0.11)	0	0	64
T3	0.05 (0.41)	0.05 (0.40)	0.43 (0.18)	0.47 (0.17)	0	0	71
T4	0.02 (0.52)	0.03 (0.52)	0.44 (0.19)	0.47 (0.19)	0	0	152
C1	1.57 (17.14)	-0.08 (1.07)	0.33 (0.26)	0.48 (0.21)	7	0	116
C2	-0.17 (3.19)	0.00 (0.26)	0.44 (0.26)	0.49 (0.11)	4	0	65
C3	0.24 (3.66)	-0.06 (0.37)	0.31 (0.26)	0.50 (0.17)	9	0	201
C4	0.06 (1.41)	0.03 (0.90)	0.32 (0.27)	0.47 (0.22)	10	0	158
C5	0.01 (1.25)	-0.07 (0.83)	0.22 (0.22)	0.32 (0.20)	13	2	131
C6	1.03 (10.23)	0.02 (0.81)	0.58 (0.28)	0.76 (0.13)	3	0	109
C7	-0.09 (0.92)	-0.02 (0.68)	0.28 (0.22)	0.44 (0.18)	10	0	113
C8	-0.42 (4.96)	-0.06 (1.47)	0.46 (0.27)	0.53 (0.21)	4	0	115
CBA	-0.06 (0.74)	-0.06 (0.53)	0.35 (0.26)	0.48 (0.21)	5	0	118
CBE	-0.23 (3.10)	-0.13 (2.33)	0.33 (0.26)	0.48 (0.20)	6	0	144
C9	1.11 (0.90)	1.04 (0.29)	0.23 (0.17)	0.30 (0.13)	6	0	51
C10	1.13 (1.18)	1.02 (0.32)	0.55 (0.23)	0.63 (0.13)	3	0	51

für ein derartiges Beispiel).

- (vi) Insgesamt ist Robustifizierung in keinem Fall nachteilig, aber häufig sehr von Vorteil gewesen. Dies bezieht sich insbesondere auch auf die i.a. geringere Variabilität der Schätzwerte. Eine robuste Behandlung ist daher - zumindest als Teilanalyse - nachdrücklich zu empfehlen.
- (vii) Die praktische Berechnung stieß in den betrachteten Fällen auf keinerlei Probleme. Lediglich im Lauf C3 wurde einmal die Iterationsschranke 201 erreicht; die dann aktuellen Werte $\hat{\mu}^{(201)}=1.70$ und $\hat{\rho}^{(201)}=0.95$ bedeuten gleichwohl gegenüber den Startwerten $\hat{\mu}=3.77$ und $\hat{\rho}=0.92$ eine Verbesserung. Die Rechenzeiten lagen zwischen 6.145 Sekunden (für C10) und 12.629 Sekunden (für T1). Dies ist sehr gering, zumal jeder Lauf 100 Replikationen umfaßt.
- (viii) Die Robustifizierung hat offenbar einen günstigen Effekt auf die Häufigkeit negativer Schätzwerte für die Varianzkomponente σ_a^2 . Es handelt sich dabei eigentlich nicht um ein Problem der robusten Statistik, da auch bei Vorliegen von reinen Normalverteilungen negative Schätzwerte auftreten, insbesondere falls σ_a^2 und damit ρ oder $\gamma := -\frac{\sigma_a^2}{\sigma_e^2}$ nahe bei 0 liegen. Nach den Tabellen von Verdooren (1982, S. 347) wird man für den Fall $\sigma_a^2 = \sigma_e^2 = 1$ und $n=p=10$ keine negativen Schätzwerte erwarten. Das stimmt mit den Simulationsresultaten für die Normalverteilungsmodelle N1-N9 und die "Tukey-Kontaminationsmodelle" T1-T4 überein. Für die "Cauchy-Kontaminationsmodelle" C1-C10 treten negative Varianzkomponentenschätzungen in etwa 4-10% der Fälle auf; nicht jedoch bei den robustifizierten Schätzern. In der Situation von C5 wird man im nicht kontaminierten Fall wegen $\frac{\sigma_a^2}{\sigma_e^2} \approx 0.43$ ungefähr 2 bis 3 negative Schätzwerte erwarten; dies entspricht in der Tat ungefähr dem Wert von RANEG. Sollen negative Schätzungen gänzlich ausgeschlossen werden, kann man z.B. den nichtnegativen Hartung-Schätzer benutzen und robustifizieren. Alternativ kann man σ_a^2 auch als Korrelation ansehen (Verdooren 1982, Smith/Murray 1984). Auch die Robustifizierung solcher Modelle ist analog

möglich; der geometrische Hintergrund wurde in 4.3.6. erwähnt.

- (ix) Schließlich kann man fragen, wie sich eine unkontaminierte, aber von der Normalverteilung verschiedene "realistische" Verteilung auswirken würde. Als ein Beispiel wurde sowohl für VA wie auch für VE die t_3 -Verteilung benutzt (bei $\mu=0$ und $n=p=10$). Es ergaben sich die Werte $MU=0.1$ (0.57) und $RMU=0.08$ (0.52) bzw. $RHO=0.40$ (0.21) und $RRHO=0.46$ (0.18). Wie erwartet war auch hier die robustifizierte Version überlegen.

4.3.13. Bemerkung: Außer der in 4.3.7. beschriebenen Methode sind weitere Ansätze zur Robustifizierung der Ein-Weg-Klassifikation denkbar. Eine erste Möglichkeit besteht darin, parallel zur klassischen ANOVA-Tafel durch gestaffelte eindimensionale Robustifizierungen eine robuste ANOVA durchzuführen. Im klassischen Fall berechnet man für $i=1, \dots, n$

den Mittelwert der $Y_{ij} =: \bar{Y}_{i.}$

die Varianz der $Y_{ij} =: V_i^2$

und hieraus

den Mittelwert der $\bar{Y}_{i.} =: \bar{Y}_{..} \longrightarrow \mu$

die Varianz der $\bar{Y}_{i.} = \frac{1}{n-1} \sum_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 \longrightarrow \frac{1}{p} \cdot [p \cdot \sigma_a^2 + \sigma_e^2]$

den Mittelwert der $V_i^2 = \frac{1}{n} \sum_i V_i^2 \longrightarrow \sigma_e^2$

(Hinter den Pfeilen stehen jeweils die hierdurch geschätzten Größen im Modell.) Analog könnte man für jedes $i=1, \dots, n$ zunächst robust

die Lokation der $Y_{ij} =: m_i$

den Skalenparameter der $Y_{ij} =: s_i$

schätzen und hieraus robust

die Lokation der $m_i \longrightarrow \mu$

den Skalenparameter der $m_i \longrightarrow \sqrt{\frac{1}{p} (p \cdot \sigma_a^2 + \sigma_e^2)}$

die Lokation der s_1^2

→ σ_e^2

ermitteln. Der Vorteil dieser Methode ist ihre Einfachheit; der Nachteil liegt darin, daß die multivariate Struktur von (7) nicht berücksichtigt wird. Falls etwa p sehr klein ist (im Extremfall etwa 2), so läßt sich eine robuste Schätzung von m_1 und s_1 nicht mehr durchführen. Auch sind die entstehenden Größen eigentlich nur im Normalverteilungsfall wirklich im Sinne des ANOVA-Modells interpretierbar. Als zweite Möglichkeit kann man (bei gegebenem $\mu, \sigma_a^2, \sigma_e^2$) durch eine Transformation die Zufallsvariablen Y_{ij} unkorreliert und im Normalverteilungsfall somit unabhängig machen. Eine Robustifizierung ließe sich dann im eindimensionalen auf der Basis der Residuen durchführen; dies entspräche i.w. Hubers (1981, S. 18) Vorschlag zur allgemeinen Robustifizierung des "Least-Squares-Ansatzes". Allerdings brauchen (wie bei der Regression) Ausreißer durchaus nicht durch große Residuen aufzufallen; es werden also u.U. die falschen Beobachtungen "winsorisiert". Generell stellt sich hier das Problem, ob und inwieweit Robustifizierung und vereinfachende Transformation miteinander vertauscht werden können: in 4.3.7. wurde nur die unerhebliche Transformation 4.3.6. benutzt, 4.3.10.(iii) ist eine Zwischenstufe, der obige Vorschlag ein Extremfall. Schließlich kann man das Problem dadurch angehen, daß man statt der Varianzkomponenten zu (robusten) "varianzähnlichen Funktionalen" (Rocke 1983) übergeht. Diese Funktionale (Grenzwerte der Stichprobenvarianzen von robusten Schätzern) sind jedoch konzeptionell wie mathematisch nicht leicht zu fassen und laufen technisch letztlich i.w. auf eine Kombination der gerade skizzierten zwei Möglichkeiten hinaus (Rocke 1983, S. 426). Überdies wird man i.a. bei Benutzung eines Varianzkomponentenmodells auf die substanzwissenschaftliche Bedeutung der Varianzkomponenten (oder daraus abgeleiteter Größen) nicht verzichten können. Die Benutzung der multivariaten Struktur aus 4.3.7. bringt zwei Beschränkungen mit sich:

- (i) Der Ansatz ist strenggenommen nur für balancierte Versuchspläne gültig. Für schwach unbalancierte Daten (z.B. $n = 10, p \in \{9, 10, 11\}$) kann man das Problem aber als "missing value" Problem ansehen und durch Einsetzen "passender" Werte (z.B. von fitted values) lösen. Dies scheint relativ unproblematisch, weil ja ohnehin robustifiziert wird.

- (ii) Für extrem kleine n (z.B. $n = 2$) ist Unempfindlichkeit gegenüber Ausreißern in der Verteilung der zufälligen Effekte a prinzipiell nicht erreichbar. Dies ist eine modellimmanente Schwierigkeit, die auch durch andere Robustifizierungsmethoden nicht überwunden werden kann.

4.3.14. Notiz (über weitere Anwendungen): Wir notieren zum Abschluß zwei nichttriviale Beispiele für weitere Anwendungen:

- (i) Meßtechnik: In der Meßfehleranalyse geht man häufig von einem Modell der Form

$$Y_{ij} = \alpha_j + \beta_j a_i + \varepsilon_{ij}$$

aus. Dabei bezeichnet Y_{ij} den mit der j -ten Methode ($j = 1, \dots, p$) gemessenen Wert des i -ten Objekts ($i = 1, \dots, n$) und a_i den (unbekannten) wahren Wert des i -ten Objekts. Ist $p = 2$, $a_i \sim N(0, \sigma_a^2)$ und sind beide Meßmethoden "unbiased" (d.h. $\alpha_1 = \alpha_2 = 0$ und $\beta_1 = \beta_2 = 1$) mit den Präzisionen σ_1^2 und σ_2^2 , so ergibt sich unter den üblichen Annahmen für $Y_i := (Y_{i1}, Y_{i2})^T$ ein Wiederholungsmodell 4.1.1.(1) mit

$$E(Y_i) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ und } \text{Cov}(Y_i) = \begin{bmatrix} \sigma_a^2 + \sigma_1^2 & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 + \sigma_2^2 \end{bmatrix}.$$

Dies führt genau im Fall gleicher Präzision ($\sigma_1^2 = \sigma_2^2$) auf ein Modell des Typs 4.3.3. Im Fall ungleicher Präzision kann man etwa die bekannten Grubbs-Schätzer robustifizieren.

- (ii) Statistische Genetik: An jeweils p Töchtern von n Bullen werden je 2 Merkmale gemessen. Ein häufig benutztes Modell zur Halbgeschwisteranalyse ist dann

$$Y_{ijk} = \mu_i + \alpha_{ij} + \varepsilon_{ijk}.$$

Dabei bedeuten:

Y_{ijk} : gemessene Ausprägung des Merkmals

μ_i : fixer Effekt für das Merkmal ($i = 1, 2$)

α_{ij} : zufälliger Effekt des j -ten Bullen auf das i -te Merkmal ($i = 1, 2$; $j = 1, \dots, n$)

ε_{ijk} : zufälliger Fehler.

Ferner wird angenommen:

$$\alpha_{ij} \sim N(0, \sigma_1^2) \quad \text{mit} \quad 0 < \sigma_1^2 < \infty \quad (i = 1, 2)$$

$$\varepsilon_{ijk} \sim N(0, \tau_1^2) \quad \text{mit} \quad 0 < \tau_1^2 < \infty \quad (i = 1, 2)$$

$$\text{Cov}(\alpha_{ij}, \alpha_{i\ell}) = \sigma_c \quad \text{mit} \quad -\infty < \sigma_c < \infty \quad (\text{für } i \neq \ell)$$

$$\text{Cov}(\varepsilon_{ijk}, \varepsilon_{i\ell k}) = \sigma_{ec} \quad \text{mit} \quad -\infty < \sigma_{ec} < \infty \quad (\text{für } i \neq \ell).$$

Alle weiteren Kovarianzen verschwinden; die gemeinsame Verteilung der Y_{ijk} sei normal. Das Modell läßt sich dann leicht in ein robustifizierbares Wiederholungsmodell des Typs 4.1.1. umschreiben; als unabhängige Wiederholungen fungieren dabei die (zu einem Bullen gehörigen) Halbgeschwistergruppen. (Da jeweils zwei Merkmale gemessen werden, handelt es sich um n Vektoren der Länge $2p$.) Die aus genetischer Sicht interessanten Parameter sind insbesondere:

$$\text{Heritabilität (Merkmal 1)} := \frac{4\sigma_1^1}{\tau_1^2 + \sigma_1^2}$$

$$\text{Heritabilität (Merkmal 2)} := \frac{4\sigma_2^2}{\tau_2^2 + \sigma_2^2}$$

$$\text{Genetische Korrelation} := \frac{\sigma_c}{\sqrt{\sigma_1^2 \cdot \sigma_2^2}}.$$

Wei uns der Leser auch fr unsre Krze Dank?

Wohl kaum. Denn Krze ward durch Vielheit leider! lang.

(G. E. Lessing)

5. SCHLUSSBEMERKUNGEN

Untersuchungen auf dem Gebiet der Statistik weisen neben einer eher mathematisch-technischen i.d.R. auch eine eher statistisch-interpretatorische und eine anwendungsbezogene Dimension auf. Im vorliegenden Problem stehen diese drei Ebenen in engem Bezug zu den in 1.1.7.(iii) aufgezhlten Sichtweisen und werfen vielfltige weitere Fragen auf, von denen abschlieend einige kurz genannt werden sollen.

Die mathematisch-technische Ebene ist eng mit der Sichtweise der Situation als Fixpunktproblem verknpft. Hier knnte durch weitere Analyse der kritischen Matrix von speziellen Modellen die Existenz- und Eindeutigkeitsfrage fr weitere Situationen gelst werden. Darber hinaus ist auch die algorithmische Seite von Interesse (Path-following Methoden), da bislang fr den multivariaten Fall kein brauchbarer Algorithmus zur Verfgung steht (Huber 1981, S. 237). Fr beide Problemkreise wird man (im diskreten Fall) u.U. weitere Hilfsmittel aus der nonsmooth Analysis heranziehen knnen. Alle diesbezglichen berlegungen sind als technische Hilfsmittel fr jeden Ansatz zur robusten Statistik von Interesse, nicht nur fr das im vorangegangenen analysierte Robustifizieren mit Konzentrationsellipsoiden.

Auf der statistisch-interpretatorischen Ebene kann man das Problem als Approximationsproblem (Anpassen einer Gleichverteilung) auffassen. Alternativ knnten aber auch andere Verteilungen betrachtet werden. Verwendet man z.B. statt der im vorangehenden stets benutzten Gleichverteilung eine winsorisierte Normalverteilung, so erhlt man eine direkte Verallgemeinerung von Hubers (1984, S. 96f) Proposal 2 (vgl. auch 2.5.4.(ii)). Die oben benutzte Grundidee zum Robustifizieren komplexerer Modelle (- nmlich die Verwendung von Ellipsoiden, die zu dem jeweiligen Problem "passen" -) wird dadurch nicht berhrt. Aus datenanalytischer Sicht kann man letztere Idee u.U. sogar umgekehrt fr die

Modellwahl nutzen, indem man Dempsters (1972) Vorschlag zur Wahl "einfacher Kovarianzstrukturen" zu robustifizieren versucht und dabei "robuste" Konzentrationsellipsoide (im Sinne dieser Arbeit) statt der ursprünglichen benutzt.

Die anwendungsbezogene Ebene bezieht sich auf die Verwendung der verarbeiteten Information und daher für die hier zunächst betrachteten Lokations- und Skalenprobleme auf Sichtweite b) aus 1.1.7.(iii). Für die in Kapitel 4 angesprochenen konkreten Modelle wäre unter diesem Aspekt zu untersuchen, inwieweit etwa verfügbare weitere substanzwissenschaftliche Information bzw. Fragestellungen in die Robustifizierung eingebracht werden können. Ein Beispiel ist etwa die Nichtnegativitätsproblematik für Varianzkomponenten. Außerdem gehören zu diesem Gebiet grundsätzliche entscheidungstheoretische Überlegungen wie etwa die Wahl des Informationsbegriffs, des Optimalitätsbegriffs und letztlich auch der "Grundmenge" für die ungünstigste Verteilung. Schon die sehr bruchstückhaften Anmerkungen im Abschnitt 2.6. deuten hier weitere Entwicklungsmöglichkeiten und Querverbindungen an. Schließlich zählen hierzu auch "klassische" Erweiterungen der ausschließlich auf Punktschätzung bezogenen obigen Untersuchungen. Dabei ist z.B. an eine Anwendung dieser Methoden zur Identifikation von Ausreißern oder zur Konstruktion robuster Tests gedacht.

L I T E R A T U R V E R Z E I C H N I S

- Aggarwal, N.L. (1974): Sur l'information de Fisher. In: Kampé de Fériet, J./Picard, C.F. (eds.): Théories de l'Information, Lecture Notes in Mathematics 398, Springer-Verlag, Berlin 111-117.
- Anderson, T.W. (1970): Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In: Bose, R.C./Chakravarti, I.M./Mahalanobis, P.C./Rao, C.R./Smith, K.J.C. (eds.): Essays in Probability and Statistics, The University of North Carolina Press, Chapel Hill, 1-24.
- Anderson, T.W. (1984): Estimating linear statistical relationships, Ann.Statist. 12, 1-45.
- Arnold, S.F. (1981): The Theory of Linear Models and Multivariate Analysis, Wiley, New York.
- Atzmon, A. (1975): A moment problem for positive measures on the unit disc, Pacific Journal of Mathematics 59, 317-325.
- Balasoorya, U./Chan, L.K. (1981): Robust estimation of Stigler's data sets by cross validation, Applied Statistics 30, 170-177.
- Barndorff-Nielsen, O. (1977): Exponentially decreasing distributions for the logarithm of particle size, Proc.Roy.Soc.Lond. A 353, 401-419.
- Barndorff-Nielsen, O. (1979): Models for non-Gaussian variation, with applications to turbulence, Proc.Roy.Soc.Lond. A 368, 501-520.
- Bauer, H. (1958): Minimalstellen von Funktionen und Extrempunkte, Archiv der Mathematik 9, 389-393.
- Bebbington, A.C. (1978): A method of bivariate trimming for robust estimation of the correlation coefficient, Applied Statistics 27, 221-226.
- Beckenbach, E.F./Bellman, R. (1965): Inequalities, Second Revised Printing, Springer-Verlag, Berlin.
- Bickel, P.J. (1965): On some robust estimates of location, Ann. Math.Statist. 36, 847-858.
- Bickel, P.J. (1976): Another look at robustness: A review of reviews and some new developments, Scandinavian Journal of Statistics 3, 145-168.
- Bickel, P.J./Lehmann, E.L. (1975): Descriptive statistics for non-

- parametric models: I. Introduction, *Ann.Statist.* 3, 1038-1044.
- Boos, D.D./Serfling, R.J. (1980): A note on differentials and the CLT and LIL for statistical functions, with application to M-estimates, *Ann.Statist.* 8, 618-624.
- Carroll, R.J. (1978a): On almost sure expansions for M-estimates, *Ann.Statist.* 6, 314-318.
- Carroll, R.J. (1978b): On the asymptotic distribution of multivariate M-estimates, *Journal of Multivariate Analysis* 8, 361-371.
- Cheng, P./Fu, J.C. (1986): On a fundamental optimality of the maximum likelihood estimator, *Statistics & Probability Letters* 4, 173-178.
- Chernoff, H. (1952): A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann.Math. Statist.* 23, 493-507.
- Chow, S.N./Hale, J.K. (1982): *Methods of Bifurcation Theory*, Springer-Verlag, New York.
- Clarke, B.R. (1983): Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations, *Ann.Statist.* 11, 1196-1205.
- Clarke, B.R. (1986): Nonsmooth analysis and Fréchet differentiability of M-functionals, *Prob.Th.Rel.Fields* 73, 197-209.
- Clarke, F.H. (1983): *Optimization and Nonsmooth Analysis*, Wiley, New York.
- Collins, J.R./Wiens, D.P. (1985): Minimax variance M-estimators in ϵ -contamination models, *Ann.Statist.* 13, 1078-1096.
- Cramér, H. (1946): *Mathematical Methods of Statistics*, Princeton University Press, Princeton.
- Dempster, A.P. (1972): Covariance selection, *Biometrics* 28, 157-175.
- Donoho, D.L./Huber, P.J. (1983): The notion of breakdown point. In: Bickel, P.J./Docksum, K.A./Hodges, J.L. Jr. (eds.): *A Festschrift for Erich L. Lehmann*, Wadsworth, Belmont, 157-184.
- Donsker, M.D./Varadhan, S.R.S. (1975): Asymptotic evaluation of certain Markov process expectations for large time, I, *Comm. Pure App.Math.* 28, 1-47.
- Donsker, M.D./Varadhan, S.R.S. (1976): Asymptotic evaluation of certain Markov process expectations for large time -III., *Comm.*

- Pure Appl.Math. 29, 389-461.
- Drygas, H. (1972): The estimation of residual variance in regression analysis, Math.Operationsforsch.Statist. 3, 373-388.
- Eisenack, G./Fenske, C. (1978): Fixpunkttheorie, Bibliographisches Institut, Mannheim.
- Ellis, R.S. (1985): Entropy, Large Deviations and Statistical Mechanics, Springer-Verlag, New York.
- Federer, H. (1969): Geometric Measure Theory, Springer-Verlag, Berlin.
- Ferentinos, K./Papaioannou, T. (1981): New parametric measures of information, Information and Control 51, 193-208.
- Fernholz, L.T. (1983): Von Mises Calculus for Statistical Functionals, Lecture Notes in Statistics 19, Springer-Verlag, New York.
- Fu, J.C. (1973): On a theorem of Bahadur on the rate of convergence of point estimators, Ann.Statist. 1, 745-749.
- Fu, J.C. (1975): The rate of convergence of consistent point estimators, Ann.Statist. 3, 234-240.
- Fu, J.C. (1982): Large sample point estimation: A large deviation approach, Ann.Statist. 10, 762-771.
- Fu, J.C. (1985): On exponential rates of likelihood ratio estimators for location parameters, Statistics & Probability Letters 3, 101-105.
- Fuglede, B. (1983): The multidimensional moment problem, Expositiones Mathematicae 1, 47-65.
- Gantmacher, F.R. (1986): Matrizentheorie, Springer-Verlag, Berlin.
- Garcia, C.B./Zangwill, W.I. (1981): Pathways to Solutions, Fixed Points and Equilibria, Prentice Hall, Englewood Cliffs.
- Gastwirth, J.L./Cohen, M.L. (1970): Small sample behaviour of some robust linear estimators of location, J.Amer.Statist.Ass. 65, 946-973.
- Gray, H.L./Odell, P.L. (1967): On least favorable density functions, SIAM Review 9, 715-720.
- Guttman, I. (1970): Statistical Tolerance Regions: Classical and Bayesian, Griffin, London.
- Hampel, F.R. (1971): A general qualitative definition of robust-

- ness, Ann.Math.Statist. 42, 1887-1896.
- Hampel, F.R./Ronchetti, E.M./Rousseeuw, P.J./Stahel, W.A. (1986): Robust Statistics: The Approach Based on Influence Functions, Wiley, New York.
- Hartung, J. (1981): Nonnegative minimum biased invariant estimation in variance component models, Ann.Statist. 9, 278-292.
- Harville, D.A. (1977): Maximum likelihood approaches to variance component estimation and to related problems, J.Amer.Statist. Ass. 72, 320-340.
- Hinderer, K. (1972): Grundbegriffe der Wahrscheinlichkeitstheorie, Springer-Verlag, Berlin.
- Hodges, J.L. Jr. (1967): Efficiency in normal samples and tolerance of extreme values for some estimates of location. In: Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1, 163-186.
- Huber, P.J. (1964): Robust estimation of a location parameter, Ann.Math.Statist. 35, 73-101.
- Huber, P.J. (1968): Robust confidence limits, Z.Wahrscheinlichkeitstheorie Verw.Gebiete 10, 269-278.
- Huber, P.J. (1977): Robust covariance. In: Gupta, S.S./Moore, D.S. (eds.): Statistical Decision Theory and Related Topics II, Academic-Press, New York, 165-191.
- Huber, P.J. (1981): Robust Statistics, Wiley, New York.
- Humak, K.M.S. (1984): Statistische Methoden der Modellbildung III, Akademie-Verlag, Berlin.
- Istrătescu, V.I. (1981): Fixed Point Theory, D. Reidel Publishing Company, Dordrecht.
- Jacobson, H.I. (1969): The maximum variance of restricted unimodal distributions, Ann.Math.Statist. 40, 1746-1752.
- Jaynes, E.T. (1957): Information theory and statistical mechanics, Physical Review 106, 620-630.
- Jeggle, H. (1979): Nichtlineare Funktionalanalysis, B.G. Teubner, Stuttgart.
- Kagan, A.M./Linnik, Yu.V./Rao, C.R. (1973): Characterization Problems in Mathematical Statistics, Wiley, New York.
- Karlin, S./Shapley, L.S. (1953): Geometry of Moment Spaces, Memoirs of the American Mathematical Society Number 12, American Mathematical Society, Providence.

- Karlin, S./Studden, W.J. (1966): Tschebycheff Systems: With Applications in Analysis and Statistics, Wiley, New York.
- Kester, A.D.M./Kallenberg, W.C.M. (1986): Large deviations of estimators, *Ann.Statist.* 14, 648-664.
- Kerridge, D.F. (1961): Inaccuracy and inference, *J.Roy.Statist. Soc. B* 23, 184-194.
- Klotz, J.H./Milton, R.C./Zacks, S. (1969): Mean square efficiency of estimators of variance components, *J.Amer.Statist.Ass.* 64, 1383-1402.
- Krein, M.G. (1959): The ideas of P.L. Čebyšev and A.A. Markov in the theory of limiting values of integrals and their further development, *American Mathematical Society Translations, Series 2*, 12, 1-121.
- Krein, M.G./Nudel'man, A.A. (1977): The Markov Moment Problem and Extremal Problems, *Translations of the American Mathematical Society*, Providence.
- Landau, H.J. (1987): Maximum entropy and the moment problem, *Bulletin (New Series) of the American Mathematical Society* 16, 47-77.
- Lanford, O.E. (1973): Entropy and equilibrium states in classical statistical mechanics. In: Lenard, A. (ed.): *Statistical Mechanics and Mathematical Problems*, *Lecture Notes in Physics* 20, Springer-Verlag, Berlin, 1-113.
- Lehmann, E.L. (1983): *Theory of Point Estimation*, Wiley, New York.
- Mardia, K.V./Kent, J.T./Bibby, J.M. (1979): *Multivariate Analysis*, Academic Press, London.
- Maronna, R.A. (1976): Robust M-estimators of multivariate location and scatter, *Ann.Statist.* 4, 51-67.
- Marshall, A.W./Olkin, I. (1979): *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York.
- McGregor, J.L. (1980): Solvability criteria for certain N-dimensional moment problems, *Journal of Approximation Theory* 30, 315-333.
- Mises, R. v. (1947): On the asymptotic distribution of differentiable statistical functions, *Ann.Math.Statist.* 18, 309-348.
- Mulholland, H.P./Rogers, C.A. (1958): Representation theorems for distribution functions, *Proc.Lond.Math.Soc.* (3) 8, 177-223.
- Müller, D.W. (1984): *Die Normalverteilung - Schwierigkeiten eines statistischen Modells*, Universität Heidelberg, Sonderforsch-

- Olbricht, W. (1987): On mergers of distributions and distributions with exponential tails, Fakultät und Institut für Mathematik der Ruhr-Universität Bochum, Bericht Nr. 109.
- Olkin, I. (1974): Inference for a normal population when the parameters exhibit some structure. In: Proschan, F./Serfling, R.J.(eds.): Reliability and Biometry, SIAM, Philadelphia.
- Olkin, I./Press, S.J. (1969): Testing and estimation for a circular stationary model, Ann.Math.Statist. 40, 1358-1373.
- Ortega, J.M./Rheinboldt, W.C. (1970): Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York.
- Ostrowski, A. (1973): Solutions of Equations in Euclidean and Banach Spaces, Academic Press, New York.
- Pearson, E.S. (1962): Some thoughts on statistical inference, Ann. Math.Statist. 33, 394-403.
- Pfanzagl, J. (1969): On the measurability and consistency of minimum contrast estimates, Metrika 14, 249-272.
- Pfanzagl, J. (1980): Asymptotic Expansions in Parametric Statistical Theory. In: Krishnaiah, P.R. (ed.): Developments in Statistics, Volume 3, Academic Press, New York, 1-97.
- Pfanzagl, J. (1981): Contributions to a General Asymptotic Statistical Theory, Lecture Notes in Statistics 13, Springer-Verlag, New York.
- Pukelsheim, F. (1977): Punktschätzungen in linearen statistischen Modellen, Dissertation, Albert-Ludwigs-Universität, Freiburg im Breisgau.
- Rényi, A. (1961): On measures of entropy and information. In: Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume I, University of California Press, Berkeley, 547-561.
- Rocke, D.M. (1983): Robust statistical analysis of interlaboratory studies, Biometrika 70, 421-431.
- Rosenberger, J.L./Gasko, M. (1983): Comparing location estimators: Trimmed means, medians, and trimean. In: Hoaglin, D.C./Mosteller, F./Tukey, J.W. (1983): Understanding Robust and Exploratory Data Analysis, Wiley, New York, 297-338.
- Schönholzer, H. (1979): Robuste Kovarianz, Dissertation, Eidgenössische Technische Hochschule, Zürich.
- Seaman, J.W./Odell, P.S./Young, D.M. (1985): Maximum variance unimodal distributions, Statistics & Probability Letters 3,

- Searle, S.R. (1982): *Matrix Algebra Useful for Statistics*, Wiley, New York.
- Serfling, R.J. (1980): *Approximation Theorems of Mathematical Statistics*, Wiley, New York.
- Shohat, J.A./Tamarkin, J.D. (1943): *The Problem of Moments*, Mathematical Surveys Number I, American Mathematical Society, New York.
- Sievers, G.L. (1978): Estimates of location: A large deviation comparison, *Ann.Statist.* 6, 610-618.
- Smith, D.W./Murray, L.W. (1984): An alternative to Eisenhart's model II and mixed model in the case of negative variance estimates, *J.Amer.Statist.Ass.* 79, 145-151.
- Stigler, S.M. (1977): Do robust estimators work with real data?, *Ann.Statist.* 5, 1055-1098.
- Strasser, H. (1985): Einführung in die lokale asymptotische Theorie der Statistik, Bayreuther Mathematische Schriften, Heft 19.
- Stute, W. (1986): Parameter estimation in smooth empirical processes, *Stochastic Processes and their Applications* 22, 223-244.
- Titterton, D.M. (1978): Estimation of correlation coefficients by ellipsoidal trimming, *Applied Statistics* 27, 227-234.
- van Campenhout, J./Cover, T.M. (1981): Maximum entropy and conditional probability, *IEEE Transactions on Information Theory*, Vol. IT-27, 483-489.
- Varadhan, S.R.S. (1984): *Large Deviations and Applications*, SIAM, Philadelphia.
- Verdooren, L.R. (1982): How large is the probability for the estimate of a variance component to be negative?, *Biometrical Journal* 24, 339-360.
- Weizsäcker, H. v./Winkler, G. (1980): Non-compact extremal integral representations: Some probabilistic aspects. In: Bierstedt, K.-D./Fuchssteiner, B. (eds.): *Functional Analysis: Surveys and Recent Results II*, North-Holland Publishing Company, Amsterdam, 115-148.
- Weyl, H. (1964): *Die Idee der Riemannschen Fläche*, Vierte Auflage, B.G. Teubner, Stuttgart.
- Wiener, N. (1961): *Cybernetics*, Second Edition, The M.I.T. Press, Cambridge, Massachusetts.

- Wiens, D. (1986): Minimax variance M-estimators of location in Kolmogorov neighbourhoods, Ann.Statist. 14, 724-732.
- Witting, H. (1985): Mathematische Statistik I, B.G. Teubner, Stuttgart.
- Zeidler, E. (1986): Nonlinear Functional Analysis and its Applications I, Springer-Verlag, New York.

LEBENS LAUF

1. Persönliche Daten:

Name:	Walter Olbricht
Geburtsdatum:	31. 5. 1956
Geburtsort:	Wuppertal - Elberfeld
Eltern:	Josef und Hildegard Olbricht

2. Schulbildung:

17. 4. 1963	Katholische Volksschule Reichsgrafen-
- 30. 11. 1966	straße in Wuppertal - Elberfeld
1. 12. 1966	Städtisches Gymnasium Wuppertal Süd
- 14. 6. 1975	in Wuppertal - Elberfeld
	(14. 6. 1975 Reifeprüfung)

3. Studium:

WS 1975/76	Mathematik mit Nebenfach Wirtschafts-
- SS 1977	wissenschaft an der Gesamthochschule
	Wuppertal
	(19. 10. 1977 Diplom-Vorprüfung II)
WS 1977/78	Mathematik mit Nebenfach Wirtschafts-
- SS 1981	wissenschaft an der Ruhr-Universität
	Bochum
	(11. 6. 1981 Diplomprüfung)
Academic Year	Statistik an der University of
1981/82	Sheffield (GB)
	(18. 2. 1983 Master of Science)

4. Berufstätigkeit:

5. 10. 1982	Wissenschaftlicher Mitarbeiter (Teil-
- 31. 8. 1983	zeitbeschäftigung) am Fachbereich
	Mathematik der Gesamthochschule
	Wuppertal
seit 1. 9. 1983	Wissenschaftliche Hilfskraft bzw.
	Wissenschaftlicher Mitarbeiter am
	Institut für Mathematik der
	Ruhr-Universität Bochum

